UPORABA METODE DNA ČRTNIH KOD ZA HITRO IDENTIFIKACIJO ŽUŽELK, KI POVZROČAJO ŠKODO NA RASTLINAH

Vladimir GRUJIĆ[1], Mojca ROT[2], Stanislav TRDAN [3], Tanja BOHINC[4], Tanja DREO[5]

[1,5]Nacionalni inštitut za biologijo, Oddelek za biotehnologijo in sistemsko biologijo, Ljubljana
[1]Mednarodna podiplomska šola Jožeta Štefana, Ljubljana
[2]KGZS, Kmetijsko gozdarski zavod Nova Gorica, Nova Gorica
[3,4]Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za agronomijo, Ljubljana

Standardizirana DNA črtna koda gena za podenoto 1 citokrom *c* oksidaze (COI), ki so jo prvotno opisali Hebert in sod. leta 2003, se pogosto uporablja za identifikacijo žuželk, ki so škodljivci rastlin. V projektu Q-Entry (CRP V4-2003) smo preizkusili protokol za metodo določanja črtnih kod DNA z uporabo gena COI na izbranih referenčnih vzorcih žuželk v različnih razvojnih stadijih (ličinke in odrasli), zbranih v rutinskih preiskavah ali pridobljenih iz mednarodnih sodelovanj. Uvedli smo standardiziran protokol, ki vključuje: ekstrakcijo DNA, pomnoževanje gena COI v PCR, ločevanje produktov PCR v agaroznem gelu, njihovo čiščenje in sekvenciranje po Sangerju, ki ga izvaja akreditirani ponudnik. Po strogih parametrih kontrole kakovosti smo skupno zaporedje COI za posamezne vzorce sestavili s programsko opremo Geneious Prime (2023.2.1, Dotmatics). Skupna zaporedja visoke kakovosti (z rezultatom Phred >30 za posamezno bazo) pričakovane dolžine (približno 600 bp) so bila pridobljena za vse vzorce iz referenčnega materiala. Končna identifikacija vzorcev iz referenčnega materiala je bila izvedena s primerjavo njihovih skupnih zaporedij COI z zaporedji v referenčnih podatkovnih bazah EPPO-Q-Bank, BOLD in NCBI GenBank. Z uporabo standardiziranega postopka smo z metodo določanja črtnih kod DNA uspešno identificirali vse testirane vzorce, vključno s ameriško koruzno sovko (*Spodoptera frugiperda*), pesno sovko (*S. exigua*), bombaževčevo sovko (*S. littoralis*), kapusovim belinom (*Pieris brassicae*), tobakovim resarjem (*Thrips tabaci*), paradižnikovo sovko (*Chrysodeixis chalcites*), paradižnikovo listno zavrtalko (*Liriomyza bryoniae*), cvetličnim resarjem (*Frankliniella occidentalis*) kot tudi žuželke izbranih neidentificiranih vzorcev. Projekt Q-Entry je učinkovito implementiral metodo DNA črtnih kod za identifikacijo žuželk, ki lahko podpre morfometrično identifikacijo

[1]mag., Večna pot 121, SI-1000 Ljubljana, e-pošta: vladimir.grujic@nib.si
[2]dr., Pri hrastu 18, SI-5000 Nova Gorica
[3]prof. dr., Jamnikarjeva 101, SI-1000 Ljubljana
[4]dr., prav tam
[5]dr., Večna pot 121, SI-1000 Ljubljana

in razširja identifikacijske zmogljivosti na življenjska obdobja žuželk, ki presegajo tradicionalne pristope.

Ključne besede: Sekvenciranje po Sangerju, referenčne baze, citokrom *c* oksidaza podenota I, identifikacija žuželk, kontrola kakovosti konsenzusnih sekvenc

ABSTRACT

DNA BARCODING AS A RAPID METHOD FOR THE IDENTIFICATION OF INSECT PESTS OF PLANTS

The COI gene's standardised barcode, introduced by Hebert *et al*. in 2003, is widely used to identify insect pests. In the Q-Entry project (CRP V4-2003), we tested the COI DNA barcoding protocol on reference material of insects at various life stages (larvae and adults). The samples were collected through routine surveys, obtained from international collaborations. We have developed and applied a standardised protocol that includes DNA extraction, PCR amplification of COI gene, agarose gel electrophoresis, PCR product purification and Sanger sequencing, carried out by an accredited provider. Following rigorous quality control parameters, consensus COI sequences for individual samples were generated using Geneious Prime software (2023.2.1, Dotmatics). High quality consensus sequences (with a Phred score > 30 of each base) of expected length (approximately 600 bp), were obtained for all samples from the reference material. The final identification of the samples from the reference material was carried out by comparing the consensus sequences with those present in reference databases, such as EPPO-Q-Bank, BOLD and NCBI GenBank. Using this method, we were able to successfully identify insect species for all tested samples, including fall armyworm (*Spodoptera frugiperda*), beet armyworm (*S. exigua*), cotton leaf worm (*S. littoralis*), cabbage caterpillar (*Pieris brassicae*), onion thrips (*Thrips tabaci*), golden twin-spot moth (*Chrysodeixis chalcites*), tomato leaf miner (*Liriomyza bryoniae*), western flower thrips (*Frankliniella occidentalis*) as well as selected unknown samples. The Q-Entry project has effectively implemented the DNA barcoding of insects, which supports their morphometric identification and extends identification capabilities to life stages of the insects beyond traditional approaches.

Keywords: Sanger sequencing, reference databases, cytochrome *c* oxidase subunit I, insect identification, quality control of consensus sequences

1 INTRODUCTION

In the field of entomological research, the accurate identification of insect plant pests is essential for effective agricultural management. Traditional identification methods rely heavily on the morphological characteristics of insects at different life stages, which can be time-consuming and inaccurate, especially when dealing with species with similar morphology. The DNA barcoding method can successfully address these challenges (Tahir *et al*., 2018). The method utilises the partial sequence of the cytochrome *c* oxidase subunit 1 (COI) gene to identify and differentiate insect species (Hebert *et al*., 2003). In this study, within the Q-Entry project (CRP V4-2003), we

developed and tested a COI DNA barcoding protocol, adapted from the standardised international protocol (EPPO, 2021), and used it to identify and characterise insect species at different life stages on reference material and unknown insects.

2 MATERIALS AND METHODS

Reference material and sample collection
Reference material of insects at different life stages (larvae and adults) was collected through international collaborations and/or through ongoing activities of partners of the Q-Entry project (Table 1). Four additional insect samples of unknown identity were included in the analysis: 1593/13, JM/23, 1592_1 and 1592_2 (Table 2). All samples were stored in 70% ethanol at -20°C until analysis.

DNA extraction
To remove 70% ethanol, all samples were dried on filter paper at room temperature for approximately 15 minutes. Larval bodies were cut into several segments, each weighing approximately 35 - 50 mg, according to EPPO recommendations (EPPO, 2021), except for larvae whose total weight fell within this range. Whole larvae and larval head segments were placed in 1.5 mL safe-lock tubes containing lysis buffer ATL (Qiagen) and macerated for 45 - 60 s to induce cell lysis. For small adult insects, a total of 5 - 11 insects were placed in a 1.5 mL safe-lock tube with ATL buffer and disrupted using a syringe needle. DNA was further purified using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions.

PCR amplification of the COI gene and sample preparation for Sanger sequencing
The amplification of the COI gene (approximately 658 bp amplicon size) followed a protocol adapted from the EPPO standard (EPPO, 2021), using a more reliable and higher quality polymerase. Reactions were performed using FastStart Taq DNA Polymerase, dNTPack kit (Roche). Reaction mixtures contain 2.5 µL of 10x FastStart PCR Buffer, 20 mM $MgCl_2$, 0.5 µL of 10 µM PCR grade dNTP mix, 0.5 µL of 10 µM solution of each primer (LCO1490 and HCO2198), synthetised by Integrated DNA Technologies, 10 U of FastStart Taq DNA Polymerase (Roche), 18 µL ddH_2O and 1 µL of the DNA template per each reaction. PCR amplification was performed under the following conditions, optimised for the FastStart Taq DNA Polymerase Master Mix (Roche): initial denaturation step at 95 °C for 4 minutes, followed by 5 cycles of denaturation at 94 °C for 30 seconds, primer annealing at 45°C for 30 seconds, and extension at 72°C for 1 minute. This was followed by 35 cycles of denaturation at 94 °C for 30 seconds, primer annealing at 51 °C for 1 min, and extension at 72 °C for 1 min, and final extension of 7 min at 72 °C. PCR products were resolved on a 1% agarose gel, containing 140 mL of 50x modified TAE buffer and 0.7 mg/ µL of ethidium bromide (stock solution 10 mg/µL) and purified using the Montage Gel Extraction Kit (Millipore), except for the *S. frugiperda* sample (Sf_1), where the concentration of PCR product was measured using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific) and diluted to a final concentration of 10 ng/µL. A total of 5 µL of each purified PCR product and diluted unpurified PCR product of Sf_1 were each mixed with an equal volume of each of the 10 µM primers used (LCO1490 and HCO2198) for the sequencing reaction, which was performed at Eurofins Genomics.

Trace editing and consensus sequence preparation

Sanger trace files were edited using Geneious Prime software (2023.2.1, Dotmatic). Strict quality criteria were set to obtain high quality consensus sequences. Quality trimming of the traces was performed by setting the error probability limit to 0.01 and consensus sequences were generated from the overlapping parts of the traces (double reads) with a high Phred score >30 for each base. Consensus sequences that met the quality criteria were further used for identification and compared with the sequences from the reference databases EPPO-Q-Bank (using blast search against Q-bank Arthropods sequences database), BOLD (using blast search against All Barcode Records on BOLD database) and NCBI GenBank (using megablast search against Nucleotide collection (nt) database). Exact matches (100% identity on 100% of sequence coverage) were required for unambiguous identification of samples. In addition, detailed characterisation of the Sf_1 sample was performed by analysing informative nucleotide positions within the consensus sequence. Informative nucleotides are located at specific single nucleotide polymorphism (SNP) sites within the COI sequence that provide genetic information suitable for identification and differentiation of different taxa. The sequence logo showing informative nucleotides for the identification and characterisation of *S. frugiperda* was generated using the WebLogo program (Crooks *et al*., 2004). The figure was edited in BioRender (BioRender, 2024). Sequence conservation and nucleotide frequency at each position were analysed to visualise informative nucleotides useful for identification and characterization of *S. frugiperda*.

3 RESULTS AND DISCUSSION

DNA barcoding for identification of the reference material samples

COI gene fragments from all samples tested were successfully amplified under selected PCR conditions. The consensus sequences generated met the quality criteria for all PCR products.

Using the consensus sequences of the purified PCR products and the unpurified diluted PCR product of Sf_1 (Table 1), all samples were identified by obtaining an exact match of their COI sequences to the COI sequences in the BOLD database. Using the NCBI GenBank database, an exact match was obtained for all the samples except the *L. bryoniae* sample. In the case of the EPPO-Q-Bank database, exact matches were not obtained for the sample of *P. brassicae* and two samples of *T. tabaci.* For the sample of *P. brassicae*, we did not obtain a match with any of the reference sequences of *P. brassicae*. The closest match was COI sequence of the aspen borer (*Choristoneura conflictana*) with 88.36% of an identity on 100% of a sequence coverage.

The difficulty in accurately identifying the sample of *L. bryoniae* using the NCBI GenBank, as well as samples of *P. brassicae,* and *T. tabaci* using the EPPO-Q-Bank database, can be attributed to the lack of reference data in the databases for this species. Another problem with the identification of the samples using the NCBI GenBank, could also be the presence of low-quality sequences in the NCBI GenBank database.

The results demonstrated that the BOLD database was the most accurate for identifying samples from the reference material panel, given that the necessary data for accurate identification of all samples in the reference material were deposited in the database.

Table 1: Overview of reference material samples and their identification based on COI consensus sequences. 'Yes', indicate samples which were successfully identified, and whose identity corresponded to known identity of insects of reference samples with 100% of an identity' on 100% of a sequence coverage. 'No' indicate samples for which exact match could not be obtained along with the percentage of an identity to the closest match on 100% of a sequence coverage. See text for further details.

Sample	**COI (bp)**	**EPPO-Q-Bank)**	**BOLD**	**NCBI GenBank**
Spodoptera frugiperda (Sf_1), larval head	614	Yes	Yes	Yes
Spodoptera littoralis, larvae	632	Yes	Yes	Yes
Spodoptera exigua, larvae	630	Yes	Yes	Yes
Pieris brassicae, larval head	638	No (88.36%)	Yes	Yes
Chrysodeixis chalcites, larval head	633	Yes	Yes	Yes
Frankliniella occidentalis, adults	629	Yes	Yes	Yes
Frankliniella occidentalis, adults	622	Yes	Yes	Yes
Thrips tabaci, adults	612	No (99.67%)	Yes	Yes
Thrips tabaci, adults	639	No (99.67%)	Yes	Yes
Liriomyza bryoniae, adults	640	Yes	Yes	No (99.53%)

DNA barcoding for characterisation of *Spodoptera frugiperda*

Characterisation of *S. frugiperda* COI sequences in the EPPO-Q-Bank database was done. Analysis of COI sequences has allowed that *S. frugiperda* samples from the EPPO-Q-Bank can be differentiated and categorized into four distinct types based on specific positions with informative nucleotides. Positions 295, 301, 397, and 400 were highly conserved across all *S. frugiperda* types and contained nucleotides that distinguish these samples from other *Spodoptera* species present in the database: *S. exigua, S. littoralis, S. litura,* and *S. eridania*. Further differentiation into four *S. frugiperda* types was achieved using informative nucleotides at positions 133, 232, 370, 451, 532, and 562 (Figure 1). COI sequence of *S. frugiperda* sample from the reference material (Sf_1), sampled in Senegal, shared 100% of an identity on 100% of a sequence coverage with the those of Type I samples, namely samples with EPPO-Q-Bank ID 6786, sampled in the Nearctic ecozone of Canada and reference samples with EPPO-Q-Bank IDs 6787,6788,6789,6790,6791 and 6792 sampled in Costa Rica. In Senegal, *S. frugiperda* was first reported in 2018, and the way how it was spread from its native regions of Americas is still unclear (Brévault *et al.,* 2018). The identified *S. frugiperda* types were further divided into rice-associated (sfR) and corn-associated (sfC) strains. The sfC strains, which predominantly infest maize and sorghum, are noted for their higher invasiveness and damage levels to maize. In contrast, sfR strains typically invade rice, millet, and grass species (Nagoshi and Meagher, 2022), although their presence has also been reported on maize (Maruthadurai and Ramesh, 2020; Yousaf *et al*., 2022).

This distinction is significant, as strains of sfC have also been observed to exhibit greater resistance to certain chemical insecticides and genetically modified crops designed to express the *Bacillus thuringiensis* (Bt) insecticidal proteins (Paredes-Sánchez *et al*., 2021).

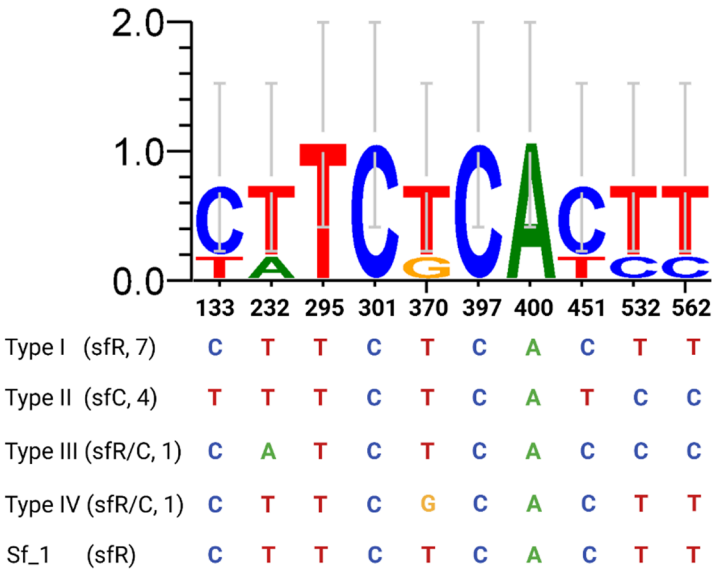

Figure 1: Sequence logo of informative nucleotides in *S. frugiperda* COI sequences. The total height of each stack, ranging from 0 to 2 bits, indicates the degree of sequence conservation at these positions, with the height of each nucleotide within a stack directly proportional to its observed frequency, illustrating its prevalence and importance in distinguishing between *S. frugiperda* types.

A comparison of COI sequences with those of reference sfR and sfC strains from the NCBI GenBank (sfR strains with accession numbers U72977.1 and KX580616.1, and sfC strain with accession number KX580614.1), revealed that all Type I samples, including Sf_1, shared identical sequences with the sfR strains, while all Type II samples shared identical sequences with sfC strains. This indicates that all Type I samples, including Sf_1, may be sfR and all Type II samples may be sfC strains. However, the potential significance of the identified informative nucleotides from Type I and Type II samples in the detection of sfR and sfC strains remains a subject for further study. The distribution of samples among the types and their strain affiliations are depicted in Figure 1, noting that Types III and IV, marked as sfR/C, suggest a divergence from the typical COI sequences of the reference sfR and sfC strains, indicating that they cannot be clearly categorized within these two strain groups.
These findings support the use of DNA barcoding not just for species identification, but also for detailed characterization, which can infer significant ecological and biological properties relevant for managing potential agricultural insect pests.

Optimisation of the tested DNA barcoding protocol

We have refined our DNA barcoding protocol by employing a reduced concentration of DNA polymerase, ranging from 1U to 5U per reaction, which has proven sufficient for the successful amplification of the COI gene fragment. This optimization was validated by analysing both the *S. frugiperda* sample from the reference material (Sf_1) and other samples in this study, including unknown specimens. The COI sequences of unpurified PCR products, diluted to a concentration of 10 ng/µL, were thoroughly examined. Our findings indicate that all generated consensus sequences met the established quality criteria. This suggests that using a lower concentration of the DNA polymerase and dilution of 10 ng/µL for unpurified PCR products is effective for insect identification and can significantly reduce the cost of analysis.

Application of DNA barcoding to unknown samples

COI gene fragments from all unknown samples were successfully amplified under defined PCR conditions. As with the reference material samples, sequences generated met the quality criteria for all PCR products.

Table 2: List of the unknown samples used in this study and identification based on their COI consensus sequences. The percentage of an identity, on 100% of a sequence coverage is given for each closest match, obtained by comparing the consensus sequences with sequences from the EPPO-Q-Bank database, the BOLD and NCBI GenBank database.

Sample	COI (bp)	(EPPO-Q-Bank)	(BOLD)	(NCBI GenBank)
1593/13 (larvae, 1)	643	*Cydalima perspectalis* (89.74%)	*Ostrinia nubilalis* (100%) *Ostrinia orientalis* (100%) *Ostrinia scapulalis* (100%)	*Ostrinia nubilalis* (100%) *Ostrinia orientalis*(100%) *Ostrinia scapulalis* (100%)
JM/23 (larvae, 1)	640	*Spodoptera litura* (89.71%)	*Agrochola macilenta* (100%)	*Agrochola macilenta* (100%)
1592_1 (adults, 5)	647	*Aphis aurantii* (90.39%)	*Rhopalosiphum maidis* (100%)	*Rhopalosiphum maidis* (100%)
1592_2 (adults, 5)	645	*Aphis aurantii* (91.31%)	*Rhopalosiphum maidis* (100%)	*Rhopalosiphum maidis* (100%)

Comparing the consensus sequences of the unknown samples with the sequences from the EPPO-Q-Bank, we did not obtain exact matches. Furthermore, their COI sequences also had relatively low percentage of an identity with sequences of their closest matches from the database (Table 2). By using BOLD and NCBI GenBank databases, samples JM/23, 1592_1 and 1592_2 were clearly identified. The exception was sample 1593/13, where we obtained exact matches for three *Ostrinia* species: European corn borer (*O. nubilalis*), Asian corn borer (*O. orientalis*) and Adzuki bean borer (*O. scapulalis*) (Table 2). These species can be grouped on the basis of their geographical distribution, with *O. nubilalis* being found in Europe and *O. orientalis* and *O. scapulalis* in Southeast Asia. However, they represent non-independent genetic units that cannot not be

distinguished based on the COI sequence (Piwczyński *et al*., 2016). We assume that sample 1593/13 is probably *O*. *nubilalis*, which is widely distributed in Slovenia.
All species of unknown samples identified using BOLD and NCBI GenBank are not listed as a quarantine pest or recommended for the regulation as quarantine pests. Due to that, the data of these insect species are not present in EPPO-Q-Bank database, which is specialized in quarantine pests, thereby explaining the lack of exact matches and low identification percentages obtained with the closest matches. However, identification of all unknown samples using BOLD and NCBI GenBank database was successful, which confirms the suitability of our implemented DNA barcoding protocol.

4 CONCLUSIONS

In this study, we report on implementation of a DNA barcoding protocol based on the COI gene, demonstrating high accuracy in identifying insect samples within the reference material. These results underscore the potential of DNA barcoding to complement traditional methods for insect identification.
Utilizing COI sequences from *Spodoptera* species in the EPPO-Q-Bank, we pinpointed 10 key positions essential for distinguishing *S. frugiperda*, categorizing them into four distinct types. Notably, the *S. frugiperda* sample from the reference material (Sf_1) was classified as Type I, sharing identical COI sequences with reference rice-associated (sfR) strains from the NCBI GenBank. This suggests that Sf_1, like other Type I samples, may be an sfR strain, which is typically manageable with chemical insecticides and *Bacillus thuringiensis* (Bt) corn, unlike the corn-associated (sfC) strains known for their resistance to these treatments.
Our protocol also successfully identified four unknown samples, indicating its applicability to other insect species that are potential plant pests.
Furthermore, our findings emphasize that the reliability of DNA barcoding significantly relies on the quality and quantity of reference data available in databases.

5 ACKNOWLEDGEMENTS

The research was funded by to the Public Agency for Research Activities of the Republic of Slovenia (ARIS) through research programmes Biotechnology and Plant Systems Biology (ARRS P4-0165) and ARRS MR Vladimir Grujić, and by the Ministry of Agriculture, Forestry and Food of the Republic of Slovenia through the project Q-Entry (V4-2003). We are grateful to Patrick Fallet, University of Neuchatel, Switzerland for samples of *Spodoptera frugiperda* and to Janja Matičič for providing sample JM/23 for our analysis.

6 REFERENCES

Brévault, T., Ndiaye, A., Badiane, D., Bal, A. B., Sembène, M., Silvie, P., Haran, J. 2018. First records of the fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae), in Senegal. Entomologia Generalis. 37: 130-42.
Crooks, G. E., Hon, G., Chandonia, J. M., Brenner, S. E. 2004. WebLogo: a sequence logo generator. Genome research. 14, 6: 1188–1190.
EPPO PM 7/129 (2). DNA barcoding as an identification tool for a number of regulated pests. 2021. EPPO Bulletin. 51, 1: 100-43.

Hebert, P. D., Cywinska, A., Ball, S. L., deWaard, J. R. 2003. Biological identifications through DNA barcodes. Proceedings. Biological sciences. 270, 1512: 313-21.

Maruthadurai, R., Ramesh, R. 2020. Occurrence, damage pattern and biology of fall armyworm, *Spodoptera frugiperda* (J.E. Smith) (Lepidoptera: Noctuidae) on fodder crops and green amaranth in Goa, India. Phytoparasitica. 48: 15–23.

Nagoshi, R.N, Meagher, R.L., 2022. The *Spodoptera frugiperda* Host Strains: What They Are and Why They Matter for Understanding and Controlling This Global Agricultural Pest. Journal of Economic Entomology. 115, 6: 1729–43.

Paredes-Sánchez, F. A., Rivera, G., Bocanegra-García, V., Martínez-Padrón, H. Y., Berrones-Morales, M., Niño-García, N., Herrera-Mayorga, V. 2021. Advances in Control Strategies against *Spodoptera frugiperda*. A Review. Molecules. 26, 18: 19 pp.

Piwczyński M., Pabijan M., Grzywacz A., Glinkowski W., Bereś PK., Buszko J. 2016. High regional genetic diversity and lack of host-specificity in *Ostrinia nubilalis* (Lepidoptera: Crambidae) as revealed by mtDNA variation. Bulletin of Entomological Research. 106, 4: 512-21.

Tahir, H. M., Noor, A., Mehmood, S., Sherawat, S. M., Qazi, M. A. 2018. Evaluating the accuracy of morphological identification of insect pests of rice crops using DNA barcoding. Mitochondrial DNA. Part B, Resources. 3, 2: 1220–24.

Yousaf, S., Rehman, A., Masood, M. Ali, K., Suleman, N. 2022. Occurrence and molecular identification of an invasive rice strain of fall armyworm *Spodoptera frugiperda* (Lepidoptera: Noctuidae) from Sindh, Pakistan, using mitochondrial cytochrome c oxidase I gene sequences. J Plant Dis Prot. 129: 71–78.