Pre-release detection of a biocontrol agent: combining independent and public DNA sequences to identify the first North American record of *Aulacidea pilosellae* (Hymenoptera: Cynipidae)

Chandra E. Moffat,[1,2] M. Alex Smith

Abstract—Here we report the first North American detection of the gall wasp *Aulacidea pilosellae* Kieffer (Hymenoptera: Cynipidae), native to Central Europe and a promising candidate biological control agent for invasive hawkweeds (*Pilosella* Vaillant, Asteraceae) in North America. This occurrence was discovered through the intersection of (i) publically available DNA barcode data and (ii) DNA sequencing of a biocontrol agent before its release. COI DNA sequences of *A. pilosellae* collected in Central Europe were compared with publically available DNA sequence records. Despite the presence of other *Aulacidea* Ashmead in the database, the most similar sequence to the European *A. pilosellae* was an as yet unidentified specimen collected in the Ottawa Valley of eastern Ontario, Canada. Subsequent sequencing of a second gene region (28S-D2) of the Ottawa Valley specimen yielded an identical DNA sequence to the European *A. pilosellae*, confirming the presence of this species in Ontario. This note highlights the potential synergy that can result from making DNA barcode data publically available before formal taxonomic identification, and a new benefit of incorporating DNA sequencing of standardised markers into biological control programmes.

Résumé—Nous rapportons la première détection nord-américaine de la guêpe à galles, *Aulacidea pilosellae* Kieffer (Hymenoptera: Cynipidae), espèce indigène en Europe centrale et candidat prometteur comme agent de lutte biologique contre les épervières (*Pilosella* Vaillant, Asteraceae) envahissants d'Amérique du Nord. La détection a été indiquée par (i) les données publiquement disponibles de séquences d'ADN pour la ressource de la communauté et (ii) le séquençage d'ADN indépendant d'un agent de lutte biologique, avant sa libération. Des séquences d'ADN de COI d'*A. pilosellae* collectés en Europe centrale ont été comparées aux données publiquement disponibles des séquences d'ADN. En dépit de la présence d'autres *Aulacidea* Ashmead dans la base de données, la séquence la plus semblable aux *A. pilosellae* européens était un spécimen jusqu'à présent non identifié collecté dans la vallée d'Ottawa dans l'est de l'Ontario (Canada). Le séquençage d'une deuxième région de gène (28S-D2) des *A. pilosellae* européens et du spécimen trouvé dans la vallée d'Ottawa a rapporté des séquences identiques. Cette note accentue la synergie potentielle qui peut résulter en rendant des données de séquences d'ADN publiquement disponibles avant identification taxonomique formelle et en incorporant le séquençage des régions normalisées d'ADN aux programmes de lutte biologique.

In this study, we document the first North American record of the gall wasp *Aulacidea pilosellae* Kieffer (Hymenoptera: Cynipidae), a species native to Central Europe, collected in the Ottawa Valley, Ontario, Canada. The detection of this species in Canada was possible due to (i) rapidly available public DNA barcode data and (ii) DNA sequencing of a biocontrol agent, before

Received 12 May 2013. Accepted 24 February 2014.

C.E. Moffat,[1,2] Department of Biology, University of British Columbia, Okanagan Campus, 3333 University Way, Kelowna, British Columbia, Canada
M.A. Smith, Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada

[1]Present address: Department of Biology, University of New Brunswick, 10 Bailey Drive/PO Box 4400, Fredericton, New Brunswick, Canada E3B 5A3.

[2]Corresponding author (e-mail: chandra.moffat@gmail.com).
Subject Editor: Chris Schmidt
doi:10.4039/tce.2014.48

its release. Here we release DNA sequences and collateral information associated with our detection of *A. pilosellae*.

In both Canada and the United States of America, *A. pilosellae* is being explored as a promising candidate biocontrol agent for several species of hawkweeds in the genus *Pilosella* Vaillant (Asteraceae), including *P. caespitosa* (Dumortier) Sell and West, *P. glomerata* (Froelich) Arvet-Touvet, and *P. piloselloides* (Villars) Soják (Grosskopf *et al.* 2008; Moffat *et al.* 2013). As part of the host range assessment for this biocontrol programme, the lead author of this paper (C.E.M.) was investigating the potential existence of morphologically cryptic types (host races or species) of *A. pilosellae*, by examining sequence variation in two gene regions, cytochrome *c* oxidase I (COI) and 28S-D2, the D2 region of the large-subunit of RNA 28S (Moffat 2012). Through a routine comparison of these *A. pilosellae* sequences to all publically available DNA sequence records on GenBank (http://www.ncbi.nlm.nih.gov/genbank/), using a standard nucleotide basic local alignment search tool (BLASTn; Altschul *et al.* 1997), the COI sequence of the European *A. pilosellae* was found to be most similar to that of a specimen identified only to order, and differing by only one base pair (a third codon synonymous C/T substitution occurring in the fourth membrane spanning helix). Surprisingly, this Hymenoptera specimen was collected in eastern Ontario, Canada (Table 1), not in Central Europe as was expected. As no prior sequence records exist for *A. pilosellae*, it was expected that the most similar sequences/species determined from the BLASTn search would be one of the four other species of *Aulacidea* (*A. freesei* Nieves-Aldrey, *A. hieracii* Bouché, *A. phlomica* Belizin, or *A. tragopogonis* Thomson) for which GenBank records for COI sequences had already been created (DQ012627, DQ012628, DQ012629, and AY368922, respectively). Instead, the COI sequence identified only as "Hymenoptera sp.", and collected in eastern Ontario, emerged as the most similar to the COI sequence of *A. pilosellae* (Table 1). While another of the European *Aulacidea* species, *A. hieracii*, has been detected in North America (Sliva and Shorthouse 2006), we are confident that our Ontario specimen is not a member of that species as it differs substantially from *A. pilosellae* in both mitochondrial and nuclear markers and in body size (Moffat 2012).

After verifying the identity and geographic origin of the Ontario specimen (done by C.E.M. contacting its collector, M.A.S.), this discussion led to the sequencing of one additional gene region for the Ontario specimen, the nuclear region 28S-D2. This revealed an identical sequence to that of the *A. pilosellae* collected in Europe (note: the 28S-D2 sequence from the Ontario specimen was identical to most specimens collected in Europe, and varied by $< 0.2\%$ from all collected specimens; (Moffat 2012).

We feel that the type of communication described here exemplifies the use of DNA barcodes and a productive use of a public DNA library – in which both pre-publication and pre-taxonomic annotation data can be useful (*i.e.*, upon discovery of the DNA sequence match, the Ontario specimen was identified only to order). While clearly there is no restriction to the use or distribution of data on GenBank (Benson *et al.* 2008), this case demonstrates that the best interests of each author were served by discussion and coordination. The publication of this note should serve as an example of the productive integration of DNA barcode data by both data generator and data communities in the development, dissemination, and integration of public data sets. As exemplified here, such collaboration led to the first documentation of a European species in North America, which may affect the *Pilosella* hawkweed biocontrol programme by potentially decreasing the screening time before introduction. To date, we know of relatively few other published cases where a non-North American species was first documented in North America using DNA barcodes (*e.g.*, deWaard *et al.* 2009; Humble *et al.* 2009; deWaard *et al.* 2010; Fernandez-Triana 2010). We believe this is the first DNA barcode-based detection of a candidate biocontrol agent before release.

Aulacidea pilosellae is a small (1.0–1.5 mm) univoltine to bivoltine (trivoltine) cynipid, known to induce small galls, most commonly on the leaf midribs, but also on the stems and stolons, on select members of hawkweeds in the genus *Pilosella* (Dalla Torre and Kieffer 1910; Eady and Quinlan 1963; Buhr 1964). *Aulacidea pilosellae* is known from Central Europe (Ionescu 1957) and

Table 1. Voucher data for specimens of *Aulacidea pilosellae* collected in Central Europe and the Ottawa Valley, Ontario, Canada.

Collection locality	Collectors	Date collected	Genbank accession	Bold accession id	Specimen depository
Wilberforce township, Ontario, Canada 45.51°N 76.977°W	M.A. Smith	14–31 May 2010	CO1: JN288739 28S-D2: KC970155	ASGLE1501-10	Biodiversity Institute of Ontario, Guelph, Ontario, Canada
Gorges du Pichoux, Switzerland 47.292°N 7.222°E	C.E. Moffat D.J. Ensing	1 July 2010	CO1: KF026457 28S-D2: KF026458	AULPL001-13	Biodiversity and Landscape Ecology Research Facility, University of British Columbia, Okanagan, British Columbia, Canada

the Mediterranean Basin (Houard 1913), and is reported from Germany (Dalla Torre and Kieffer 1910); France (Folliot 1964); Switzerland, Poland, and the Czech Republic (Moffat *et al.* 2013); Spain (Nieves-Aldrey *et al.* 2005); the United Kingdom (Eady and Quinlan 1963); Israel (Argaman 1988); Romania (Ionescu 1957); and Hungary (Sárospataki 1999). Across its native range, it is most commonly reported on *Pilosella officinarum* Vaillant (Asteraceae) (as the synonymised name *Hieracium pilosella* Linnaeus; Dalla Torre and Kieffer 1910; Ionescu 1957; Eady and Quinlan 1963), but also reported from several other *Pilosella* species (Buhr 1964, Moffat *et al.* 2013). All hawkweeds in the genus *Pilosella* present in North America are non-native and of European origin.

Collection of *A. pilosellae* in North America consisted of setting a Malaise trap (Townes 1962) erected on a small ridge near the Bonnechere River, Ontario, Canada in April 2010 and maintained until November 2010 (Table 2). The trap was emptied every two weeks and trap contents were maintained at −20 °C. A high-resolution panoramic photograph of the collection site (taken on 2 April 2010, approximately eight weeks before the trapping event) can be viewed at http://www.gigapan.com/gigapans/46381. The unsorted sample was transferred to the University of Guelph (Ontario, Canada) where contents were sorted to order and then morphospecies. Selected specimens were photographed (Fig. 1) and then tissue sampled for DNA extraction (Table 1).

Table 2. Timeline for the collection, extraction, sequencing, and querying of the sequence from the Ottawa Valley specimen in a public database.

Event	Date
Collection – Malaise trap open	14–31 May 2010
DNA extraction	11 May 2011
COI amplification	12 May 2011
COI sequencing	16 May 2011
GenBank record created for COI	15 July 2011
Initial email query from C.E.M. to M.A.S.	12 October 2011

Sequencing of the North American *A. pilosellae* consisted of a DNA extract prepared from single leg using a glass-fibre extraction protocol (Ivanova *et al.* 2006). The DNA extracts were re-suspended in 30 µl of dH_2O, and a the CO1 DNA barcode region (a 658-base pair region near the 5' terminus of the COI gene) was amplified using standard insect barcoding region primers LepF1 (5'-ATTC AACCAATCATAAAGATATTGG-3') and LepR1 (5'-TAAACTTCTGGATGTCCAAAAAATCA-3') following established protocols (as in Smith *et al.* 2008). The variable D2 region of the rDNA 28S region was amplified and sequenced using the primers D2B (5'-GTCGGGTTGCTTGAGA GTG-3') and D3Ar (5'-TCCGTGTTTCAAGA CGGGTC-3') (Saux *et al.* 2004). The resultant amplicons were bi-directionally sequenced. All laboratory information for the individual sequences can be retrieved from the Barcode of Life Data

Fig. 1. The specimen of *Aulacidea pilosellae* (Kieffer) (ASGLE2-0266) collected in late May 2010 in the Ottawa Valley, Ontario, Canada (45.51°N, 76.977°W).

System (BOLD) (Ratnasingham and Hebert 2007), using the Process ID (sequence accession). All sequence data and detailed collection information is available on BOLD (www.barcodinglife.org) in the public data set: First Canadian Record of *Aulacidea pilosellae* (Cynipidae) (dx.doi.org/10.5883/DS-ASCYN1) and in Table 1.

Collections of *A. pilosellae* in Central Europe were made by rearing specimens from galls on whole live *Pilosella* plants in June 2010, as described in Moffat *et al.* (2013). Sequencing of Central European *A. pilosellae* were conducted as described in Moffat (2012). Briefly, the primers LCOI490 (5'-GGTCAACAAATCATAAAGATATTGG-3') and HCO2198 (5'-TAAACTTCAGGGTGACCAAAAAATCA-3') (Folmer *et al.* 1994) were used to amplify the CO1 gene region, and the primers 28S-D2(F) (5'-CGTGTTGCTTGATAGTGCAGC-3') and 28S-D2(R) (5'-TCAAGACGGGTCCTGAAAGT-3') (Heraty *et al.* 2004; Ács *et al.* 2010) were used to amplify the D2 region of the nuclear 28S rDNA gene. Sequences for both regions are available in GenBank and BOLD (Table 1).

The standard nucleotide basic local alignment search tool (BLAST*n*) (Altschul *et al.* 1997) was used to search nucleotide databases (all GenBank and other databases; http://blast.ncbi.nlm.nih.gov) for the most similar available sequences to the sequenced COI gene regions. We used all default search options (*i.e.*, Database = *Others* (*Nucleotide collection* (*nr/nt*)); Optimize for: *Highly similar sequences* (*megablast*)).

The COI sequence from the Ontario collection, as well as the Central European collection (Table 1), is characteristic of Hymenoptera mitochondrial DNA with a high AT content (74%). BOLD automatically assigns globally unique identifiers (GUI) to specimens that are 2% divergent from other records in the database (for details on the derivation see Ratnasingham and Hebert 2013). These different GUI may or may not represent different species – but in either case serve as a useful "label" for specimens that do not yet have a taxonomically assigned species name. At the time of writing, searches in BOLD for this GUI (BOLD:AAU8720) yield five additional public BOLD records (JSHYM075-11, JSHYM293-11, JSHYM593-11, JSHYN760-11, JSHYP286-11) collected at a different eastern Ontario locality (north of Brockville, Ontario, Canada; 44.6214°N, 75.7734°W), indicating that our detection is not an isolated incident. These data associated with these public specimens, and others that are recovered in the future, are all retrievable from BOLD by searching the GUI BOLD:AAU8720. The CO1 DNA sequence of the Ontario specimen and associated record (Table 1) was made rapidly public following sequencing, as part of the iBOL program (ibol.org). The time delay (two months) was sufficient to perform simple validation tests regarding the ordinal identification of the sequence on the BOLD database and to rule out possible contamination, but not to assign a taxonomically valid species name (Table 2).

Part of the DNA barcoding initiative is the community-wide agreement to the policy of rapid data release pre-publication (following the Fort Lauderdale Principles – Wellcome Trust 2011). Here, in broad terms, there are "data consumers" (DC) and "data producers" (DP). Data producers agree to produce high-quality data and to make that data immediately and freely available. Data consumers agree to appropriately cite the course and acknowledge the DP's while being free to use the data in any creative way. In particular, "… the best interests of the community are served when all act responsibly to promote the highest standards of respect for the scientific contribution of others. In some cases, this might best be done by discussion or coordination with the resource producers" (Wellcome Trust 2011).

By detecting the presence of *A. pilosellae* in Canada, the process for approving this candidate

biocontrol agent for release in North America may potentially be expedited. Regardless of the outcome of this particular case, there is clear potential significance of both (i) the DNA barcoding of candidate biocontrol agents and (ii) making publically available standardised DNA sequences even from un-identified specimens. While DNA sequencing of candidate biological control agents is becoming more common (Gaskin *et al.* 2011), in large part to confirm host associations and investigate cryptic genetic variation, we advocate a protocol whereby standardised DNA regions (such as the CO1 DNA barcode) of candidate species are not only sequenced, but then routinely compared with available sequences in databases such as GenBank and BOLD. Such a protocol may reduce the amount of time required to petition candidate agents, should the species already have established but gone undocumented in the target country.

Acknowledgements

The authors would like to thank C. Smith and E. Bazinet Smith for diligently monitoring the Ottawa Valley Malaise trap where the Canadian wasp was collected in May 2010. They thank P. Coghlin and K. Floate, Agriculture and Agri-Food Canada (AAFC), Lethbridge, for sequencing the European *A. pilosellae* and identifying its similarity to the Ontario specimen. They also thank R. Lalonde, J. Pither, G. Grosskopf-Lachat, and R. De Clerck-Floate for providing the opportunity to conduct the cryptic diversity work on *A. pilosellae*. They thank J. Mlynarek for assisting with translation of the abstract. M.A.S.'s work is supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Leaders Opportunity Fund (LOF) from the Canada Foundation for Innovation (CFI). Laboratory analyses of the COI sequence generated from the North American collection were funded by the Government of Canada through Genome Canada and the Ontario Genomics Institute (2008-0GI-ICI-03). C.E.M. was supported by a NSERC Canada Graduate Scholarship and by NSERC Discovery Grant funds to R. Lalonde and J. Pither. Financial support for the sequencing work was provided by AAFC Research Project #2642 (R. De Clerck-Floate) and a NSERC Discovery Grant (R. Lalonde).

References

Ács, Z., Challis, R.J., Bihari, P., Blaxter, M., Hayward, A., Melika, G., *et al.* 2010. Phylogeny and DNA barcoding of inquiline oak gallwasps (Hymenoptera: Cynipidae) of the western Palaearctic. Molecular Phylogenetics and Evolution, **55**: 210–225.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., *et al.* 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research, **25**: 3389–3402.

Argaman, Q. 1988. Additions to the cynipoid fauna of Israel (Hymenoptera, Cynipoidea). Israel Journal of Entomology, **22**: 109–118.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. 2008. GenBank. Nucleic Acids Research, **36**: D25–D30.

Buhr, H. 1964. Bestimmungstabellen der Gallen (Zoo- und Phytocecidien) an Pflanzen Mittel- und Nordeuropas. Veb Gustav Fischer Verlag, Jena, Germany.

Dalla Torre, K.W. and Kieffer, J.J. 1910. Cynipidae. *In* Das Tierreich. *Edited by* F.E. Schultze. Liefer, Berlin, Germany. Pp. 1–891.

deWaard, J.R., Humble, L.M., and Schmidt, B.C.S. 2010. DNA barcoding identifies the first North American records of the Eurasian moth, *Eupithecia pusillata* (Lepidoptera: Geometridae). Journal of the Entomological Society of British Columbia, **107**: 25–31.

deWaard, J.R., Landry, J.F., Schmidt, B.C., Derhousoff, J., McLean, J.A., and Humble, L.M. 2009. In the dark in a large urban park: DNA barcodes illuminate cryptic and introduced moth species. Biodiversity and Conservation, **18**: 3825–3839.

Eady, R.D. and Quinlan, J. 1963. Hymenoptera: Cynipoidea. Key to families and subfamilies and Cynipinae (including galls). Royal Entomological Society of London, London, United Kingdom.

Fernandez-Triana, J. 2010. Eight new species and an annotated checklist of Microgastrinae (Hymenoptera: Braconidae) from Canada and Alaska. ZooKeys, **63**: 1–53. doi:10.3897/zookeys.63.565.

Folliot, R. 1964. Contribution à l'étude de la biologie des Cynipides gallicoles (Hymenopteres: Cynipoidea). Annales des Sciences Naturelles Zoologie, **12**: 407–564.

Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology, **3**: 294–299.

Gaskin, J.F., Bon, M.-C., Cock, M.J.W., Cristofaro, M., Biase, A.D., De Clerck-Floate, R., *et al.* 2011. Applying molecular-based approaches to classical biological control of weeds. Biological Control, **58**: 1–21.

Grosskopf, G., Wilson, L.M., and Littlefield, J.L. 2008. Host-range investigations of potential biological control agents of alien invasive hawkweeds (*Hieracium* spp.) in the USA and Canada: an overview. *In* Proceedings of the XII international symposium on biological control of weeds. *Edited by* M.H. Julien, R. Sforza, M.-C. Bon, H.C. Evans, P.E. Hatcher, H.L. Hinz, and B.G. Rector. CAB International, Wallingford, United Kingdom. Pp. 552–557.

Heraty, J.M., Hawks, D., Kostecki, J.S., and Carmichael, A. 2004. Phylogeny and behavior of the Gollumiellinae, a new subfamily of the ant parasitic Eucharitidae (Hymenoptera: Chalcidoidea). Systematic Entomology, **29**: 544–559.

Houard, C. 1913. Les Zoocécidies des plantes d'europe et du Bassin de la Méditerranée. Libraire Scientifique A. Hermann et Fils, Paris, France.

Humble, L.M., deWaard, J.R., and Quinn, M. 2009. Delayed recognition of the European poplar shoot borer, *Gypsonoma aceriana* (Duponchel) (Lepidoptera: Tortricidae) in Canada. Journal of the Entomological Society of British Columbia, **106**: 61–70.

Ionescu, M.A. 1957. Cynipinae. Fauna Republicii Populare Romine, Volume IX, Editura Academiei Republicii Populare Romine, Bucharest, Romania.

Ivanova, N.V., deWaard, J.R., and Hebert, P.D.N. 2006. An inexpensive, automation-friendly protocol for recovering high-quality DNA. Molecular Ecology Notes, **6**: 998–1002.

Moffat, C.E. 2012. Cryptic host-associated and frequency-dependent patterns of host species selection of a candidate weed biological control agent in its native range. Masters thesis. University of British Columbia. Available from http://hdl.handle.net/2429/43199 [accessed 4 June 2014].

Moffat, C.E., Lalonde, R.G., Ensing, D.J., De Clerck-Floate, R.A., Grosskopf-Lachat, G., and Pither, J. 2013. Frequency-dependent host species use by a candidate biological control insect within its native European range. Biological Control, **67**: 498–508.

Nieves-Aldrey, J.-L., Vardal, H., and Ronquist, F. 2005. Comparative morphology of terminal-instar larvae of Cynipoidea: phylogenetic implications. Zoologica Scripta, **34**: 15–36.

Ratnasingham, S. and Hebert, P.D.N. 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). Molecular Ecology Notes, **7**: 355–364.

Ratnasingham, S. and Hebert, P.D.N. 2013. A DNA-based registry for all animal species: the barcode index number (BIN) system. Public Library of Science One, **8**: e66213.

Sárospataki, M. 1999. Phytophagous insects associated with *Hieracium pilosella* (Asteraceae) in Hungary, Central Europe. Environmental Entomology, **28**: 1–8.

Saux, C., Fisher, B.L., and Spicer, G.S. 2004. Dracula ant phylogeny as inferred by nuclear 28S rDNA sequences and implications for ant systematics (Hymenoptera: Formicidae: Amblyoponinae). Molecular Phylogenetics and Evolution, **33**: 457–468.

Sliva, M.D. and Shorthouse, J.D. 2006. Comparison of the development of stem galls induced by *Aulacidea hieracii* (Hymenoptera: Cynipidae) on hawkweed and by *Diplolepis spinosa* (Hymenoptera: Cynipidae) on rose. Canadian Journal of Botany, **84**: 1052–1074.

Smith, M.A., Rodriguez, J.J., Whitfield, J.B., Deans, A.R., Janzen, D.H., Hallwachs, W., *et al*. 2008. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proceedings of the National Academy of Sciences, **105**: 12359–12364.

Townes, H. 1962. Design for a Malaise trap. Proceeding of the Entomological Society Washington, **64**: 253–262.

Wellcome Trust 2011. Sharing data from large-scale biological research projects: a system of tripartite responsibility [online]. Report of a meeting organized by the Wellcome Trust, 14–15 January 2003, Fort Lauderdale, Florida, United States of America. Available from http://www.genome.gov/Pages/Research/WellcomeReport0303.pdf [accessed 1 May 2013.