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) publicly 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 publicly 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

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

2Corresponding 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 collater al 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* species of hawkweeds in the genus promising candidate biocontrol agent for several *A. pilosellae* sequence of the European America, collateral information associated with our detection of a specimen released. Here we release DNA sequences and

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

© 2014 Entomological Society of 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 1. Voucher data for specimens of *Aulacidea pilosellae* collected in Central Europe and the Ottawa Valley, Ontario, Canada.

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

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

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

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 dH₂O, 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'-ATTCAACCAAATCATAAGATTTTG-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'-GTGCGGTTGGTTGAGAGTTG-3') and D3Ar (5'-TCGGGTGTTTCAAGA CGGGTTC-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...
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'-CGTGTTGCTTGACTGCAGC-3') and 28S-D2(R) (5'-TCAAGACGGTTCTCTGAAAGT-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 (BLASTn) (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 COI 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 P. Coghlin and K. Floate, Agriculture and Agri-Food Canada (AAFC), Lethbridge, for sequencing 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. Mlyнarek 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


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.


© 2014 Entomological Society of Canada