Stormer, H.G., Pike, E.M., Sperling, F.A.H., and Proctor, H.C. 2025. Distribution and diversity of terrestrial isopods (Isopoda: Oniscidea) in Canada, including new records and a species checklist. The Canadian Entomologist.

## **Supplementary material, File S1. Molecular identification of sowbugs**

### **Methods**

We extracted sowbug DNA with a DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions, but with the following modifications: (1) tissue was ground with a pestle after adding buffer ATL; (2) incubation time was 3 hours after adding proteinase K; (3) 4 μL of RNAse A was added after incubation. For polymerase chain reaction (PCR) and sequencing we used the primers LCO1490 (5'‑GGTCAACAAATCATAAAGATATTGG-3') and either HC02198 (5'‑TAAACTTCAGGGTGACCAAAAAATCA-3') or IsoCoiRint (5'‑GCYCCYGCYAAWACAGGKARDGA-3'; Folmer *et al*. 1994; Dimitriou and Sfenthourakis 2022). The latter primer is designed specifically for sowbugs and usually resulted in successful amplification of samples that produced a poor sequence with the LCO1490–HC02198 primer pair. Polymerase chain reactions contained 5 μL of template DNA, 0.5 μL of each primer (10 μM concentration), 10 μL of AccuStart II GelTrack® PCR SuperMix (Quantabio, Beverly, Massachusetts, United States of America), and 4 μL of water, for a total volume of 20 μL. When using the LCO1490–HC02198 primer pair, the amplification conditions consisted of initial denaturation at 94 ˚C for 3 minutes, followed by 35 cycles of 94 ˚C for 45 seconds, 48 ˚C for 45 seconds, and 72 ˚C at 60 seconds, and a final extension at 72 ˚C for 7 minutes. Amplification with LCO1490 and IsoCoiRint followed the “strict” protocol in Dimitriou and Sfenthourakis (2022). We assessed amplification success with gel electrophoresis, cleaned amplified DNA with a column-based PCR cleanup kit (Truin Science KTS1115; Truin Science, Edmonton, Alberta, Canada) and submitted the samples to the University of Alberta Molecular Biology Service Unit (Edmonton, Alberta) for bidirectional Sanger sequencing. To create consensus sequences for each sample we manually trimmed and aligned the forward and reverse reads in JalView (Waterhouse *et al*. 2009) and searched for matching sequences in both BOLD (Ratnasingham and Hebert 2007) and NCBI BLAST (Altschul *et al*. 1990), using default settings. All sowbug *CO*1 sequences obtained in this study are deposited in GenBank under accession numbers PP995851–PP995894.

To assess confidence in species identifications resulting from *CO*1 sequencing, we constructed a maximum likelihood phylogeny for each sowbug family from which we had obtained sequences. Each phylogeny contained all *CO*1 sequences from this study and from GenBank (https://www.ncbi.nlm.nih.gov/nucleotide/) that belonged to that family. The outgroups for these phylogenies consisted of the longest *CO*1 sequences available from GenBank for every species-level taxon in Ligiidae and Tylidae, with the phylogenies rooted on Ligiidae. We did not include members of the genus *Ligia* Fabricius, 1798 (Isopoda: Ligiidae) among the outgroups because *Ligia*’s inclusion in Oniscidea is debated (Dimitriou *et al*. 2019; Tabacaru and Giurginca 2021). We aligned and trimmed the sequences for each family using MAFFT (Nakamura *et al*. 2018) and constructed each phylogeny with IQ-TREE, version 2.2.2.7 (Nguyen *et al*. 2015). We used ModelFinder (Kalyaanamoorthy *et al*. 2017) to test substitution models and choose the model with the lowest Bayesian information criterion (BIC) score for each phylogeny (-m MFP). Phylogenies were assessed with ultrafast bootstrap (Hoang *et al*. 2018) using 5000 replicates (-B 5000) and allowing up to 5000 bootstrap iterations (-nm 5000). Finally, we used FigTree (Rambaut 2009) and TreeGraph2 (Stöver and Müller 2010) to format and annotate the phylogenies for viewing.

### **Results**

We obtained 41 *CO*1 barcode sequences from 12 sowbug species representing seven families, including specimens from Alberta, Saskatchewan, British Columbia, and Ontario (Table S1). All sequences except six (2× *Oniscus asellus*, 3× *Nagurus cristatus*, 1× *Porcellionides pruinosus*) matched a species-level record from BOLD or GenBank with greater than 99% identity. For all sequences except those from *Nagurus cristatus* and one *Oniscus asellus* specimen, the closest sequence match was to the species determined by morphological identification. For the three *N.* *cristatus* sequences, the closest match in GenBank was *Hemimysis* *margalefi* Alcaraz *et al*., 1986 (Mysida: Mysidae) (78.8–79.3% identity), and the closest match in BOLD was *Mongoloniscus vannamei* (Arcangeli, 1927) (Isopoda: Agnaridae) (80.03–80.13% identity); identification of the specimens as *N. cristatus* was therefore entirely based on morphological characteristics (Stormer and Proctor 2022)*.* Please see the section below (DNA-based identification of *Nagurus cristatus*) for an investigation of the taxonomic position of *N. cristatus* and the possibility that the *N. cristatus CO*1 sequences represent nuclear mitochondrial pseudogenes (NUMTs). The two *Oniscus* *asellus* sequences matched *O. asellus* in BOLD with 82.53–94.58% identity, but in GenBank they matched different species: one of the sequences matched to *O. asellus* with 96.3% identity, and the other matched *Armadillidium vulgare* with 81.3% identity. The *Porcellionides* *pruinosus* sequence matched with 97.5% identity to *P. pruinosus* in GenBank and 97.7% identity to *P. pruinosus* in BOLD. The poor matches for *O. asellus* and *P. pruinosus* were likely due to a high level of sequence noise that resulted in many unresolved bases (Supplementary material, Fig. S2).

In the *CO*1 phylogenies constructed for each sowbug family, almost all sequences for each species of interest (including those obtained in this study) were placed in a single monophyletic clade with greater than 95% ultrafast bootstrap support (*e.g*., Supplementary material, Fig. S3E). One of two exceptions was *Nagurus cristatus*. No other sequences were available for this species besides the ones obtained in this study, which were placed in a polytomy with Tylidae and the rest of Trachelipodidae rather than clustering with *CO*1 from *Nagurus carinatus* (Dollfus, 1905) (Isopoda: Trachelipodidae)(Supplementary material, Fig. S3F)*.* The other exception was *Porcellionides pruinosus,* which was paraphyletic due to *Porcellionides myrmecophilus* (Stein, 1859) (Isopoda: Porcellionidae) forming a well-supported clade within the *P. pruinosus* clade (Supplementary material, Fig. S3E). It was not possible to assess whether *Cylisticus convexus* could be differentiated from other Cylisticidae using *CO*1 as no other cylisticids had publicly available *CO*1 sequences.

## **DNA-based identification of *Nagurus cristatus***

### **General approach**

The *CO*1 sequences we obtained in this study from *Nagurus cristatus* (Dollfus) (Isopoda: Trachelipodidae) were unexpectedly dissimilar to those from other species of Trachelipodidae, leading us to suspect that they may represent nuclear mitochondrial pseudogenes (NUMTs) rather than functional *CO*1 sequences. In this supplementary file, we describe our investigation into this possibility to highlight this molecular ambiguity and to promote subsequent study.

If the *CO*1 sequences from *N. cristatus* represent NUMTs, they would likely contain many nonfunctional substitutions; alternatively, if they represent functional *CO*1 sequences, *N. cristatus* may be highly diverged from other Trachelipodidae. We explored the possibility that the sequences are NUMTs through: (1) a comparison of our *N. cristatus CO*1 sequences with those from isopods and related crustaceans; (2) a search for alternate *CO*1 sequences in our *N. cristatus* specimens; and (3) an investigation into the possibility that *N. cristatus* is molecularly diverged from other Trachelipodidae by sequencing an additional gene, 18*S*.

### **Cytochrome *c* oxidase subunit 1 comparison**

To look for evidence of nonfunctionality, we added the three *CO*1 sequences we had obtained from *Nagurus cristatus* to a dataset of 25 additional *CO*1 sequences from GenBank consisting of 19 sowbug species (terrestrial isopods) and six nonterrestrial isopod species. We included three nonisopod peracarid outgroups consisting of two species of amphipod and the mysid *Hemimysis* *margalefi* Alcaraz *et al*. (Mysida: Mysidae) (GenBank HE614276.1) and rooted the phylogenies on *H. margalefi*. We manually aligned and trimmed the dataset in JalView (Waterhouse *et al*. 2009) and constructed a maximum likelihood tree using MEGA6 (Tamura *et al*. 2013). We then used MEGA6 to convert the nucleotide sequences to amino acid sequences, construct a pairwise distance matrix from the amino acid sequences, and build a neighbour-joining tree using the pairwise distance matrix. We also used the pairwise distance matrix to calculate the average amino acid dissimilarity for each sequence included in the matrix.

Excluding unresolved bases, all *Nagurus cristatus* sequences were identical. When analysed together with a variety of terrestrial and aquatic isopods and other Peracarida, the *N. cristatus* sequences formed a clade between the *H. margalefi* outgroup and the rest of Peracarida rather than appearing in the clade containing all other sowbugs (Supplementary material, Fig. S4). In the phylogeny constructed with only Trachelipodidae *CO*1 sequences, *Nagurus cristatus* was placed in a polytomy with the Tylidae outgroup and the rest of Trachelipodidae rather than clustering with the only other congeneric species represented in GenBank, *Nagurus carinatus* (Dollfus) (Supplementary material, Fig. S3F). The *N. cristatus* sequences did not exhibit a long branch compared to the other sowbugs in either of the phylogenies.

Nuclear mitochondrial pseudogenes may contain stop codons, but not always (Moulton *et al*. 2010). Conversion of the nucleotide sequences to amino acid sequences did not reveal stop codons in *N. cristatus* or in any of the other peracarid sequences in the dataset. The neighbour-joining tree constructed by pairwise comparison of amino acid sequences showed evidence of sequence divergence for *N. cristatus*, which had a slightly longer branch than other Trachelipodidae included in the tree (Supplementary material, Fig. S5). *Nagurus cristatus* also had the highest average amino acid sequence dissimilarity (0.136 amino acid substitutions per site) of any sowbug species in the dataset (Supplementary material, Fig. S6). The second-highest average dissimilarity was 0.125 substitutions per site for *Protracheoniscus pokarzhevskii* Gongalsky and Turbanov(Isopoda: Agnaridae) (GenBank MH400723.1).

### **Search for alternate *CO*1**

To attempt to obtain the “true” *Nagurus* *cristatus* *CO*1 sequence, we used Poly Peak Parser (Hill *et al*. 2014) to subtract the possible NUMT sequence from the mixed templates that resulted from our attempts to sequence the *CO*1 barcode region from *N. cristatus* using LCO1490 and IsoCO1Rint primers (see main paper). We also constructed two pairs of *Nagurus*-specific *CO*1 primers with Primer3 (Ye *et al*. 2012) as implemented in Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) using a *CO*1 sequence from *Nagurus* *carinatus* (Dollfus) (Isopoda: Trachelipodidae) (GenBank ON212542.1) as the template. These primer pairs are Nag\_3\_F/Nag\_3\_R (5'‑GGGGCTCTTGGGACTTCATT-3', 5'‑TAGCCCCGGCTAAAACAGGA-3') and Nag\_9\_F/Nag\_9\_R (5'‑CGGAGTTAGGTCAAGCTGGT-3', 5'‑TAGCCCCGGCTAAAACAGG-3'). The PCR volume and components were the same as described previously for *CO*1 sequencing from the other sowbug species. Amplification conditions for the *Nagurus* primers Nag\_3\_F/Nag\_3\_R were intended to limit nonspecific amplification and consisted of denaturation at 94 ˚C for 3 minutes, followed by 5 cycles of 94 ˚C for 45 seconds, 62 ˚C for 45 seconds, and 72 ˚C for 60 seconds; 5 cycles of 94 ˚C for 45 seconds, 60 ˚C for 45 seconds, and 72 ˚C for 60 seconds; 10 cycles of 94 ˚C for 45 seconds, 58 ˚C for 45 seconds, and 72 ˚C for 60 seconds; 10 cycles of 94 ˚C for 45 seconds, 55 ˚C for 45 seconds, and 72 ˚C for 60 seconds; and a final extension at 72 ˚C for 7 minutes. For the Nag\_9\_F/Nag\_9\_R primers, we used an alternative amplification protocol consisting of denaturation at 94 ˚C for 3 minutes, followed by 15 cycles of 94 ˚C for 45 seconds, 55 ˚C for 45 seconds, 72 ˚C for 60 seconds; 15 cycles of 94 ˚C for 45 seconds, 52 ˚C for 45 seconds, 72 ˚C for 60 seconds; and a final extension at 72 ˚C for 7 minutes. We assessed amplification success as described previously in our main paper.

Our *Nagurus cristatus* *CO*1 sequences contained a high level of noise when produced with the LCO1490–HC02198 primer pair, suggesting that two or more templates were present. Amplification with LCO1490–IsoCoiRint resulted in an unreadable sequence with even more noise. Elevated noise may result from coamplification of a “true” mitochondrial *CO*1 barcode and a nuclear mitochondrial pseudogene (NUMT; Moulton *et al*. 2010). Our attempts to extract the “true” barcode from the mixed template using Poly Peak Parser (Hill *et al*. 2014) did not result in a clear sequence.

### **Sowbug 18*S* phylogeny**

If the *CO*1 sequences obtained from *Nagurus cristatus* represent the “true” *CO*1 sequence, this would suggest that *N.* *cristatus* is extremely diverged from other sowbugs, which should be reflected in its nuclear genome. We used the primers 18SA1\_mod (5'‑CTGGTTGATCCTGCCAGTCATATGC-3'; Raupach *et al*. 2009) and 18S700R (5'‑CGCGGCTGCTGGCACCAGAC-3'; Dreyer and Wägele 2001) to amplify a 620-bp section of the 18*S* gene from the same three *N. cristatus* specimens used for *CO*1 sequencing. Amplification conditions followed a touchdown PCR protocol developed by Dimitriou *et al*. (2018). We assessed amplification success and obtained final sequences as described previously. *Nagurus cristatus* partial 18*S* sequences obtained in this study are deposited in GenBank under accession numbers PQ001581–PQ001583.

To construct a phylogeny to compare partial 18*S* sequences in *Nagurus cristatus* and other sowbugs, we compiled a dataset of 18*S* sequences by searching GenBank (https://www.ncbi.nlm.nih.gov/nucleotide/) for “18*S* and FAMILY” for all 39 extant sowbug families (Ahyong *et al*. 2011; Javidkar *et al*. 2015; Sfenthourakis and Taiti 2015) and added the longest sequence for each species to the dataset. We also added the longest sequence for each genus that did not have any representatives identified to species level. We aligned these sequences with MAFFT (Nakamura *et al*. 2018) using the E-INS-I alignment algorithm, which is recommended for RNA polymerase sequences (https://mafft.cbrc.jp/alignment/software/algorithms/algorithms.html) and constructed a maximum likelihood phylogeny with IQ-TREE (Nguyen *et al*. 2015) using the process described previously for the sowbug *CO*1 sequences. We also constructed a second phylogeny using an alternative alignment produced with the MAFFT L-INS-I alignment algorithm. Both trees were rooted on Ligiidae. As for the *CO*1 phylogenies, we did not include *Ligia* as an outgroup (see main paper).

All three *Nagurus cristatus* specimens yielded identical 18*S* sequences excluding unresolved bases. The closest GenBank match to the consensus of the three sequences was *Philoscia muscorum* (Scopoli)(Isopoda: Philosciidae) (GenBank AJ287058.1) which matched with 91.03% identity. The *N. cristatus* 18*S* sequences were not included in the main Trachelipodidae cluster in the 18*S* phylogeny as expected, but instead were contained in a clade with Scyphacidae, Porcellionidae, and *Levantoniscus* Cardoso *et al*. (Isopoda: Trachelipodidae) (Supplementary material, Fig. S7). There were two other unexpected placements of taxa traditionally included in the Trachelipodidae: (1) *Levantoniscus* clustered with Porcellionidae, and (2) *Porcellium* Dahl (Isopoda: Trachelipodidae) was placed in Cylisticidae. Four other unexpected placements included the following: (1) *Tylos punctatus* Holmes and Gay (Isopoda: Tylidae) placed with Ligiidae instead of with *Tylos europaeus* Arcangeli (Isopoda: Tylidae);   
(2) two Philosciidae spp. placed with Platyarthridae; (3) two species of *Spherillo* Dana (Isopoda: Armadillidae) placed with *Tylos* instead of in Armadillidae; (4) *Trichorhina tomentosa* (Budde-Lund) (Isopoda: Platyarthridae) placed as an outgroup to Paraplatyarthridae, Philosciidae, and Armadillidae, rather than in Platyarthridae. The placement of the *T. punctatus* sequenceis likely due to poor sequence alignment, as the *Tylos* sequences did not align well with the rest of the dataset. Using an alternative alignment algorithm (L-INS-I) resulted in both *Tylos* sequences clustering together with *Spherillo* (Supplementary material, Fig. S8).

### **Discussion**

Although most sowbug species examined in this study yielded *CO*1 barcodes that matched closely with those from the species to which they belonged morphologically, this was not the case for *Nagurus cristatus*. Cytochrome *c* oxidase subunit 1 barcodes from this species were dissimilar from those of congeneric and confamilial relatives when compared both as nucleotide sequences and as amino acid sequences, and the nucleotide sequences contained a lot of noise. These characteristics are associated with sequences that are nuclear mitochondrial pseudogenes (NUMTs; Moulton *et al*. 2010). If so, this would mean that the *CO*1 sequences obtained in this study are not the “true” sequences for this species, which would represent the first record of an isopod NUMT, as NUMTs have not been recorded from this group (Raupach *et al.* 2022). However, the sequences contained no stop codons that would render the sequences nonfunctional, and the unusual placement of *N. cristatus* relative to other sowbug families based on its partial 18*S* sequence suggests that this species is in fact considerably diverged from other Trachelipodidae. Therefore, the *CO*1 sequences obtained in this study may actually represent the “true” sequences for this species, despite their dissimilarity to those of other Trachelipodidae. *Wolbachia*, an endosymbiotic maternally-inherited bacterium, could be a possible mechanism behind this divergence. *Wolbachia* is present in many species of arthropods, including sowbugs (Zimmermann *et al.* 2015). *Nagurus cristatus* is parthenogenetic (Gregory 2014), and *Wolbachia* infection has been linked to the rapid genetic divergence of parthenogenic weevil populations, with a correlation between *Wolbachia* infection and parthenogenesis (Elias-Costa *et al*. 2019). In any case, the divergent 18*S* sequences obtained from *Nagurus cristatus* raise the question of whether this species belongs in Trachelipodidae. Determining whether it belongs in Scyphacidae (as suggested by the 18*S* phylogeny), in a new family, or whether it is simply an unusual member of Trachelipodidae would require detailed morphological examination and further molecular investigation that are beyond the scope of this study.

Several of the other unusual placements of taxa in the 18*S* phylogeny are consistent with earlier studies. *Levantoniscus* was originally described as Trachelipodidae (Cardoso *et al*. 2015) but was placed in Porcellionidae in our phylogeny (Supplementary material, Fig. S7). The same placement also occurred in a study that included sequences from several genes in addition to 18*S* (Dimitriou *et al*. 2018), and those authors suggested that the genus might belong to a new family rather than to Trachelipodidae or Porcellionidae. Similarly, in our phylogeny, *Porcellium* was placed in Cylisticidae instead of in Trachelipodidae, and this placement also occurred in another 18*S*-based phylogeny (Mattern 2003), in which it was suggested that *Porcellium* does not belong to Trachelipodidae. The three other “misplacements” observed in the phylogeny (Philosciidae spp. with Platyarthridae; *Spherillo* spp. with *Tylos* instead of with Armadillidae; *Trichorhina tomentosa* outside Platyarthridae) are difficult to explain, as the sequences involved were not noticeably shorter, less well-aligned, or less resolved than the other sequences.

# **References**

Please see the main paper, “Distribution and diversity of terrestrial isopods (Isopoda: Oniscidea) in Canada, including new records and a species checklist”, for the references cited in this supplementary material file to that paper.

Supplementary figures to this file: Supplementary material, Figure S2, S3, S4, S5, S6, S7, S8; Table S1