Untangling taxonomy: a DNA barcode reference library for Canadian spiders

GERGIN A. BLAGOEV, JEREMY R. DEWAARD, SUJEEVAN RATNASINGHAM, STEPHANIE L. DEWAARD, LIUQIONG LU, JAMES ROBERTSON, ANGELA C. TELFER and PAUL D. N. HEBERT
Biodiversity Institute of Ontario, University of Guelph, Guelph, ON, Canada

Abstract

Approximately 1460 species of spiders have been reported from Canada, 3% of the global fauna. This study provides a DNA barcode reference library for 1018 of these species based upon the analysis of more than 30 000 specimens. The sequence results show a clear barcode gap in most cases with a mean intraspecific divergence of 0.78% vs. a minimum nearest-neighbour (NN) distance averaging 7.85%. The sequences were assigned to 1359 Barcode index numbers (BINs) with 1344 of these BINs composed of specimens belonging to a single currently recognized species. There was a perfect correspondence between BIN membership and a known species in 795 cases, while another 197 species were assigned to two or more BINs (556 in total). A few other species (26) were involved in BIN merges or in a combination of merges and splits. There was only a weak relationship between the number of specimens analysed for a species and its BIN count. However, three species were clear outliers with their specimens being placed in 11–22 BINs. Although all BIN splits need further study to clarify the taxonomic status of the entities involved, DNA barcodes discriminated 98% of the 1018 species. The present survey conservatively revealed 16 species new to science, 52 species new to Canada and major range extensions for 426 species. However, if most BIN splits detected in this study reflect cryptic taxa, the true species count for Canadian spiders could be 30–50% higher than currently recognized.

Keywords: Araneae, Barcode Index Numbers, DNA barcoding, operational taxonomic units, species identification, spiders

Received 14 March 2015; revision received 30 June 2015; accepted 6 July 2015

Introduction

With 45 000 known species in 114 families, spiders are one of the most diverse orders of arthropods, a fact which stimulated the development of an online taxonomic catalogue (World Spider Catalog 2014). It now tracks the description of every new species, providing an authoritative resource for valid names. However, it does not simplify their identification. In common with other groups of arthropods, many species of spiders are difficult to identify, especially the juveniles that often dominate collections from seasonal environments (Foelix 2011). In fact, because most keys are only effective for adults and often just for males, many individuals cannot be identified to a species level through morphological analysis. Because of this difficulty, biotic surveys frequently neglect spiders, despite their diversity and important role as predators in terrestrial ecosystems (Freetias et al. 2013).

Studies over the past decade have revealed the efficacy of DNA barcoding, the analysis of sequence diversity in the 5' region of the cytochrome c oxidase I gene, as a tool for the identification of animal species (Hebert et al. 2003). Insects have been the target for much of this work, especially members of the order Lepidoptera. These analyses have shown that 95–100% of the species in regional faunas can be discriminated (Hajibabaei et al. 2006; Hebert et al. 2009; Hausmann et al. 2011). Recent work has tested the impact on barcode resolution of expanding from a regional to continental scale. Barcodes discriminated all 75 Australian species in the family Sphingidae regardless of their collection site (Rougerie et al. 2014), while 1000 species of Lepidoptera showed only a small reduction in identification success in comparisons involving Austria and Finland (Huemer et al. 2014). Further studies have confirmed that DNA barcoding is effective in other insect orders. For example, the analysis of 1500 Finnish and 3500 German species of Coleoptera revealed 99% success in their identification (Hendrich et al. 2014; Pentinsaari et al. 2014).
The effectiveness of DNA barcoding as a tool for the identification of spiders has seen less evaluation. Barrett and Hebert (Barrett & Hebert 2005) found that DNA barcoding discriminated all 168 species of Canadian spiders that they examined, but their sample sizes were small. A subsequent study that analysed nearly 3000 specimens from a site in the Canadian Arctic indicated that each of the 198 species at this locality possessed a diagnostic array of barcode sequences (Blagoev et al. 2013). Similar results were obtained for 63 species of Alaskan spiders (Slowik & Blagoev 2012) and 361 species of Canadian spiders (Blagoev et al. 2009). Although these studies have established that DNA barcodes are very effective in identifying known species, they also revealed possible overlooked species, as evidenced by many cases of deep intraspecific divergence (Barrett & Hebert 2005; Blagoev et al. 2009). The spiders in other regions have seen less investigation. Astrin et al. (Astrin et al. 2006) reported the strong performance of DNA barcoding in identifying 52 European species in the family Pholcidae, while Miller et al. (2013a) found that barcodes discriminated 31 species in the Netherlands. Čandeš and Kuntner (2014) established that the effectiveness of DNA barcodes for identification was not impeded by geographic variation in a study on 20 species of spiders whose ranges span Europe and North America. These overall results have led to consensus that a comprehensive DNA barcode reference library will advance spider biology in important ways. For example, Miller et al. (2013b) showed how it could provide new insights into the parasitoids which attack spiders. A well-parameterized library will also facilitate the use of next-generation sequencing platforms to ascertain the species composition of mass collections (Hajibabaei et al. 2011; Yu et al. 2012; Ji et al. 2013). Because the latter advance will enable the inclusion of spiders in large-scale biotic surveys, while the former will reveal new details on their ecological interactions (Wirta et al. 2014), there is a need for further library construction.

This study reports progress towards the assembly of a DNA barcode reference library with 10x or higher coverage for all Canadian spiders. It represents a major taxonomic and geographic extension of earlier work which targeted a limited number of species and only examined them at a few localities. The goal of 10x coverage was established to permit the analysis of intraspecific variation in barcode sequences. Prior taxonomic studies have indicated the magnitude of the challenge – about 1460 species of spiders are known from Canada (Dondale 1979; Bennett 1999; Paquin et al. 2010; Dupérré 2013). These taxa include representatives from two of the three suborders and from 46 of the 114 families of spiders. Although this study does not deliver a complete reference library for Canadian spiders, it does provide an average of 20x coverage for two-thirds of the fauna, including nearly all common species. As a result, it creates a highly functional identification system for this biotic assemblage. The study has also provided a better understanding of the importance of cryptic diversity in Canadian spiders by ascertaining the incidence of species which are assigned to two or more BINs (Ratnasingham & Hebert 2013) and presents an inexpensive pathway that can speed the registration of new species.

Materials and methods

Specimen collections

Work began with the collection of spiders from sites across Canada. In contrast to many other groups of terrestrial arthropods where museum specimens are held dry on pins allowing barcode recovery for several decades (Hebert et al. 2013), spiders are typically stored in 70–80% alcohol at room temperature, conditions which lead to rapid DNA degradation (Vink et al. 2005). As a result, barcode recovery is difficult from specimens more than 5 years old, especially small-bodied species (Miller et al. 2013a). Because most of its spiders were collected prior to 2000, the Canadian National Collection of Insects, Arachnids and Nematodes (Ottawa) contributed few records to this study. However, collections at the Royal BC Museum (Victoria) and the Lyman Entomological Museum (Montreal) provided 8.3% and 0.7% of the specimens examined, reflecting their recent involvement in large-scale surveys. The remaining specimens (91%) were acquired through sampling programs led by the Biodiversity Institute of Ontario that employed diverse methods including hand collecting, sieving, sweep netting and trapping (Malaise, pan, pitfall, sticky) at sites across Canada (Fig. 1).

Specimen identifications

All specimens were identified or verified by GAB through the genitalic examination of all adults for each species when they were available. Determination often occurred following barcode analysis, once the specimen was assigned a BIN (see below). In this way, DNA barcoding not only identifies specimens that already exist in the reference library, but it also limits the downstream work of specialists by sorting the remaining specimens into operational taxonomic units that require examination (deWaard et al. 2009). Furthermore, it often guides the specialist in the determination, providing higher taxonomic assignments and phylogenetic affinities. Morphological identifications were based on a comprehensive set of reference publications on the taxonomy
of North American spiders with species nomenclature following the World Spider Catalog (2014). Collection data and a taxonomic assignment for each specimen and often a photograph are available in the Barcode of Life Data Systems (BOLD, www.boldsystems.org) (Ratnasingham & Hebert 2007) in the public data set, ‘Spiders of Canada’ at DS-SOC2014 (doi:10.5883/DS-SOC2014).

### DNA sequence analysis

A total of 30,679 specimens were sequenced (Table 1). The interval between collection and sequencing ranged from a few days to 35 years, but most specimens were analysed within 2 years ($\bar{x} = 414$ days, $\sigma = 377$). Tissue lysis, DNA extraction, PCR amplification, cycle sequencing and sequence analysis were performed at the Canadian Centre for DNA Barcoding (CCDB; www.ccdb.ca) employing standard protocols (Ivanova et al. 2006; deWaard et al. 2008; Hebert et al. 2013). For most samples, the primer cocktails of C_LepFolF and C_LepFolR (Folmer et al. 1994; Hebert et al. 2004) were used for both PCR and bidirectional sequencing of the barcode region. The residual DNA extracts are stored in the DNA archive at the CCDB where they are available for additional study. Sequences, electropherograms and primer details for each specimen are accessible on BOLD (doi:10.5883/DS-SOC2014) and GenBank (see Table S1, Supporting information for accession numbers).

### Barcode index numbers

After their upload to BOLD, all sequences >500 bp were assigned a BIN, an interim taxonomic system which aggregates similar barcode sequences into a BIN (Ratnasingham & Hebert 2013). BOLD currently contains over 375K BINs (December 2014) generated using the Refined Single Linkage (RESL) algorithm which employs a three-phased analysis to reach decisions on the number of BINs (= OTUs) in the overall sequence data set on BOLD (Ratnasingham & Hebert 2013). In contrast to some other approaches employed for OTU designation, such as Automatic Barcode Gap Discovery (Puillandre et al. 2012), its outcome is deterministic. It is also much faster than other approaches, such as the generalized mixed Yule-coalescent model (Pons et al. 2006; Fujisawa & Barracough 2013), a critical requirement for the analysis of large data sets (see Ratnasingham & Hebert 2013 for details of algorithm details and comparisons).

Although past studies have not employed postprocessing of BIN results, we assessed the robustness of each of the 1359 BINs detected in this study in the following fashion. The overall sequence alignment was exported from BOLD and partitioned phylogenetically into two alignments of approximately equal size, each including all members for a particular set of families. A NJ tree for each alignment was generated in MEGA5 (Tamura et al. 2011) using the K2P distance model, and bootstrap values for each node were determined based on 100

---

© 2015 The Authors. Molecular Ecology Resources Published by John Wiley & Sons Ltd.
Table 1  Number of species in the 46 families of spiders known from Canada and a summary of barcode coverage and recovery success

| Family               | Common name          | Feeding strategy† | Body size‡ | Length (mm) | Canadian species | % Barcoded | Specimens barcoded | COI obtained | Success rate % |
|----------------------|----------------------|-------------------|------------|-------------|------------------|------------|-------------------|--------------|----------------|
| Agelenidae           | Funnel weavers       | WEB L             | 4.0–18     | 21          | 71               | 530        | 467               | 88           |                |
| Amaurobiidae         | Hackledmesh weavers  | AMP L             | 1.3–20     | 14          | 93               | 379        | 369               | 97           |                |
| Amphinectidae        | Amphinectid spiders  | AMP M             | 6.0–10     | 1           | –                | –          | –                 | –            |                |
| Anapidae             | Ground orbweavers    | WEB M             | 1.0–1.5    | 1           | 100              | 1          | 1                 | 100          |                |
| Antrodiaetidae       | Foldingdoor spiders  | AMP L             | 6.0–25     | 4           | 75               | 25         | 20                | 80           |                |
| Anyphaenidae         | Ghost spiders        | AMP M             | 3.0–8.5    | 6           | 100              | 289        | 277               | 96           |                |
| Araneidae            | Orbweavers           | WEB L             | 1.5–30     | 64          | 89               | 2202       | 2060              | 94           |                |
| Atypidae             | Purseweb spiders     | AMP L             | 8.0–27     | 1           | 100              | 9          | 9                 | 100          |                |
| Clubionidae          | Sac spiders          | AMP M             | 3.0–10     | 33          | 76               | 1441       | 1389              | 96           |                |
| Corinniidae          | Antimimic spiders    | AMP M             | 1.0–15     | 8           | 63               | 35         | 26                | 74           |                |
| Cybaeidae            | Soft spiders         | WEB M             | 1.4–14     | 14          | 71               | 107        | 78                | 73           |                |
| Dictynidae           | Meshweavers          | WEB M             | 1.2–15     | 77          | 60               | 1499       | 1415              | 94           |                |
| Dipluridae           | Funnelweb spiders    | WEB M             | 3.5–17     | 1           | –                | –          | –                 | –            |                |
| Dysderidae           | Cell spiders         | ACP M             | 9.0–15     | 1           | 100              | 6          | 6                 | 100          |                |
| Eutichuridae         | Prowling spiders     | AMP M             | 2.8–12     | 3           | 100              | 181        | 181               | 100          |                |
| Gnaphosidae          | Stealthy ground spiders | AMP M         | 2.0–17     | 93          | 52               | 753        | 639               | 85           |                |
| Hahniidae            | Comb-tailed spiders  | WEB M             | 1.2–7.8   | 16          | 75               | 226        | 192               | 85           |                |
| Linyphiidae          | Dwarf and sheetwork weavers | WEB S          | 1.0–4.0  | 564        | 68               | 10531      | 8863              | 84           |                |
| Liocranidae          | Spinylegged sac spiders | AMP M        | 2.0–10    | 2           | 100              | 82         | 71                | 87           |                |
| Lycosidae            | Wolf spiders         | ACP L             | 2.2–35     | 104         | 80               | 3879       | 3394              | 87           |                |
| Mecistidae           | Midget funnelweb tarantulas | WEB M       | 3.5–18    | 1           | 100              | 1          | 1                 | 100          |                |
| Mimetidae            | Pirate spiders       | WEB M             | 2.5–7.0    | 8           | 75               | 50         | 42                | 84           |                |
| Miturgidae           | Prowling spiders     | AMP M             | 2.8–6.8    | 1           | 100              | 5          | 3                 | 60           |                |
| Myxosoma             | Dwarf cobweb weavers | WEB S             | 0.5–2.0    | 3           | 33               | 6          | 6                 | 100          |                |
| Nesticidae           | Cave cobweb spiders  | AMP M             | 1.8–10     | 3           | 67               | 4          | 3                 | 75           |                |
| Oecobiidae           | Flatmesh weavers     | WEB S             | 0.9–3.2    | 1           | –                | –          | –                 | –            |                |
| Oonopidae            | Goblin spiders       | AMP S             | 1.0–3.0    | 2           | –                | –          | –                 | –            |                |
| Oxyopidae            | Lynx spiders         | ACP M             | 3.2–8.2    | 2           | 100              | 18         | 18                | 100          |                |
| Philodromidae        | Running crab spiders | ACP M             | 3.0–8.2    | 48          | 65               | 1240       | 1175              | 95           |                |
| Pholcidae            | Cellar spiders       | WEB M             | 1.0–10     | 4           | 75               | 100        | 48                | 48           |                |
| Pheucticoides        | –                    | AMP M             | 1.0–15     | 15          | 47               | 213        | 200               | 94           |                |
| Pimeliidae           | Large hammockweb spiders | WEB M        | 5.0–12    | 3           | 100              | 15         | 14                | 93           |                |
| Pisauridae           | Nursery web spiders  | AMP L             | 5.0–37     | 7           | 100              | 121        | 88                | 73           |                |
| Salticidae           | Jumping spiders      | ACP M             | 1.0–22     | 115         | 65               | 1463       | 1336              | 91           |                |
| Scytodidae           | Spitting spiders     | ACP M             | 3.5–10     | 2           | 50               | 1          | 1                 | 100          |                |
| Segestriidae         | Tubeweb spiders      | WEB M             | 4.0–10     | 1           | 100              | 3          | 3                 | 100          |                |
| Sicariidae           | Recluse spiders      | AMP M             | 5.0–13     | 1           | –                | –          | –                 | –            |                |
| Telemidae            | Longlegged cave spiders | WEB S        | 1.0–1.7    | 2           | 100              | 23         | 17                | 74           |                |
| Tetragnathidae       | Longjawed orbweavers | WEB L             | 1.2–25     | 25          | 92               | 1630       | 1464              | 90           |                |
| Therididae           | Cobweb weavers       | WEB M             | 0.8–12     | 102         | 75               | 1940       | 1830              | 94           |                |
| Theridiosomatidae    | Ray orbweavers       | WEB S             | 1.3–2.4    | 1           | 100              | 26         | 23                | 88           |                |
| Thomisidae           | Crab spiders         | AMP M             | 1.5–11.3   | 66          | 77               | 1539       | 1465              | 95           |                |
| Titanidae            | Rock weavers         | AMP M             | 4.0–8.0    | 4           | 75               | 38         | 26                | 68           |                |
| Trachelidae          | –                    | AMP M             | 5.5–7.7    | 3           | 67               | 40         | 26                | 65           |                |
| Uloboridae           | Hackled orbweavers   | WEB M             | 4.0–8.0    | 5           | 80               | 28         | 23                | 82           |                |
| Zodariidae           | Ant spiders          | AMP M             | 2.0–14     | 1           | –                | –          | –                 | –            |                |
| Total                | –                    | –                 | 1454       | 70          | 30 679           | 27 269     | 89                |              |                |

†WEB – web builders (20), AMP – ambush predators (19), ACP – active predators (6).
‡L – large (9), M – medium (31), S – small (6).
replicates. With BINs mapped on the NJ tree (using ‘groups’ in MEGA5), each BIN was assigned the bootstrap value for the terminal branch leading to its members. In the few cases where a BIN was paraphyletic, a value of N/A was recorded. BINs with bootstrap values greater than 60% were viewed as well supported, while those with lower values were not. Although this postprocessing was particularly directed towards evaluating the significance of the species involved in BIN splits, all species were analysed.

Geographic variation in sequences

Analysis of geographic variation was restricted to the 339 species that were represented by at least 10 specimens collected from sites at least 1000 km apart based on the shortest path connecting all sampling coordinates using the minimum spanning tree algorithm (Prim 1957). The correlation between the distance matrices for geographical distance and sequence divergence was determined for each species using the Mantel test (Mantel 1967). Species lacking a significant correlation (P > 0.05) were excluded from subsequent analysis. The slope of the relationship between sequence divergence and geographic distance per kilometre was calculated separately for each species with a significant correlation, and this was used to estimate the rise in sequence divergence per 1000 km. For example, if the slope of the relationship between sequence divergence and distance was 0.002% per kilometre, this would indicate a 2% increase in sequence divergence per 1000 km.

Results

Sequence analysis

Barcode sequences were recovered from 89% (27 269/30 679) of the specimens that were analysed (Fig. S1, Supporting information). Although the primer sets were effective across the order, there was variation in amplification success among families ($\chi^2 = 248.4; P < 0.001$) with relatively low recovery from species in two families (Cybaeidae, Pholcidae). There was no evidence of nuclear copies of mitochondrial DNA (NUMTs), and the incidence of nontarget sequences from Proteobacteria, such as Wolbachia, was trivial as noted previously in spiders (Smith et al. 2012). Largely because of differing success in specimen acquisition, taxon coverage varied among the 46 families of spiders (Table 1). Six families lack records, but they collectively include just seven species, while the other 39 possess coverage ranging from 33 to 100% ($\bar{x} = 67\%$) of their Canadian species.

The 27 269 sequences comprised 16 622 distinct haplotypes; most haplotypes were only represented once, but 3441 were recovered from multiple individuals (range = 2–91). Intraspecific divergences averaged 0.78%, while mean distance to the NN taxon was 10-fold higher, averaging 7.85% (Table 2). As a consequence, there was a clear barcode gap for most species (Fig. 2).

BIN analysis

The BIN count was substantially higher than the species count (1359 vs. 1018), but there was only a weak relationship ($R^2 = 0.04, P < 0.001$) between the BIN count for a species and the number of specimens analysed (Fig. 3), after the three outlier species were removed. The BIN/species ratio for the 40 families varied from 0.93 to 2.74. Although it is well established that the proportion of OTUs that cannot be assigned to a named species increases for smaller-bodied taxa, there was no evidence that the B/S ratio was linked to size as small-bodied families showed a ratio similar to that for large-bodied families (Fig. 4a). The B/S ratio was significantly higher for families whose component species had broad vs. narrow distributions, but the coefficient of determination was just 16% (Fig. 4b).

There was a perfect correspondence between the specimens assigned to 795 of the BINs and the members of a particular species. The remainder of the BINs included a few cases where members of two or more species shared a BIN (merge) and many more cases where the members of a species were assigned to more than one BIN (split).

BIN merges

BIN merges were uncommon as just 15 BINs included representatives of two or more species. These cases involved 34 species, all morphologically similar congeners. Twelve of the 15 BINs included a pair of species, two others a triad, while the final case involved four species. Barcode diversities (ranging from 0.46% to 1.30%) permitted discrimination of the species in six of these pairs and one triad (Table 3). Species in the other six pairs and the other triad could not be discriminated because their component taxa shared barcodes. Three of the four species in the tetrad could be separated, but two species shared barcodes. As a consequence, 998 of the 1018 species (98%) possessed a unique array of barcode sequences permitting their identification, while the others could be assigned to a species pair or, in one case, a triplet.

BIN splits

BIN splits were far more common than merges, being detected in 197 species (Table S2, Supporting information). Maximum sequence divergences between the BINs assigned to a currently recognized species averaged
4.32%, but ranged from 1.24% to 12.11%. Many species represented by multiple specimens were placed in either a single BIN or a few, but three (Grammonota angusta Dondale 1959, Tetragnatha versicolor Walckenaer 1841, Evarcha hoyi Peckham & Peckham 1883) were outliers, being assigned to 22, 20 and 11 BINs, respectively (Table S2, Supporting information; Figs 3 and 5). Most of the BINs had bootstrap values above 60%, and no linkage was apparent between bootstrap values and number of specimens analysed (Fig. 6a). There was a positive

Table 2 Measures of barcode diversity in 1018 species of Canadian spiders. BIN splits indicate the number of species in each family whose members were assigned to two or more BINs. BIN merges indicate the number of BINs that included members of two or more species.

| Family        | Species barcoded | BINs | Mean intraspecific divergence | Mean NN distance based on species | BIN splits | BIN merges |
|---------------|------------------|------|-------------------------------|----------------------------------|------------|------------|
| Agelenidae    | 15               | 14   | 0.27                          | 6.58                              | –          | 1          |
| Amaurobiidae  | 13               | 13   | 0.23                          | 9.54                              | –          | –          |
| Amphinectidae | –                | –    | –                             | –                                 | –          | –          |
| Anapidae      | 1                | 1    | –                             | 16.39                             | –          | –          |
| Antheridiidae | 3                | 5    | 2.51                          | 13.01                             | 1          | –          |
| Anypheanidae  | 6                | 6    | 0.48                          | 9.00                              | –          | –          |
| Araneidae     | 57               | 72   | 0.66                          | 9.03                              | 8          | –          |
| Atypidae      | 1                | 1    | 0.26                          | 11.82                             | –          | –          |
| Clubionidae   | 25               | 27   | 0.57                          | 7.02                              | 2          | –          |
| Corinnidae    | 5                | 5    | 0.43                          | 8.62                              | –          | –          |
| Cybaeidae     | 10               | 12   | 0.97                          | 8.02                              | 2          | –          |
| Dictynidae    | 46               | 81   | 1.14                          | 8.61                              | 13         | –          |
| Dipluridae    | –                | –    | –                             | –                                 | –          | –          |
| Dysderidae    | 1                | 2    | 1.14                          | 16.01                             | 1          | –          |
| Eutichuridae  | 3                | 4    | 0.31                          | 11.33                             | 1          | –          |
| Gnaphosidae   | 48               | 55   | 0.68                          | 7.64                              | 4          | –          |
| Hahniidae     | 12               | 21   | 0.93                          | 10.98                             | 5          | –          |
| Linyphiidae   | 384              | 510  | 0.73                          | 7.82                              | 71         | 5          |
| Liocranidae   | 2                | 3    | 1.00                          | 4.23                              | 1          | –          |
| Lycosidae     | 83               | 96   | 0.46                          | 4.16                              | 15         | 6          |
| Mecicobothriidae | 1           | 1    | –                             | 15.79                             | –          | –          |
| Mimetidae     | 6                | 6    | 0.39                          | 9.02                              | –          | –          |
| Miturgidae    | 1                | 1    | 0.00                          | 7.54                              | –          | –          |
| Mysmenidae    | 1                | 1    | 0.16                          | 13.00                             | –          | –          |
| Nesticidae    | 2                | 2    | 0.00                          | 19.89                             | –          | –          |
| Oecobiidae    | –                | –    | –                             | –                                 | –          | –          |
| Oonopidae     | –                | –    | –                             | –                                 | –          | –          |
| Oxyopidae     | 2                | 3    | 2.32                          | 10.04                             | 1          | –          |
| Philodromidae | 31               | 46   | 1.23                          | 8.56                              | 11         | 1          |
| Pholcidae     | 3                | 4    | 2.04                          | 22.78                             | 1          | –          |
| Phrurolithidae| 7                | 9    | 0.64                          | 7.91                              | 2          | –          |
| Pimoidae      | 3                | 4    | 1.45                          | 13.62                             | 1          | –          |
| Pisauridae    | 7                | 7    | 0.50                          | 4.36                              | –          | –          |
| Salticidae    | 75               | 110  | 1.18                          | 7.57                              | 24         | 2          |
| Scytodidae    | 1                | 1    | –                             | 3.05                              | –          | –          |
| Segestriidae  | 1                | 1    | 0.51                          | 17.1                              | –          | –          |
| Sicariidae    | –                | –    | –                             | –                                 | –          | –          |
| Telemidae     | 2                | 2    | 0.62                          | 15.65                             | –          | –          |
| Tetragnathidae| 23               | 61   | 1.34                          | 12.52                             | 9          | –          |
| Theridiidae   | 76               | 90   | 0.67                          | 7.93                              | 9          | –          |
| Theridiosomatida | 1            | 1    | 0.82                          | 10.43                             | –          | –          |
| Thomisidae    | 51               | 71   | 1.12                          | 5.92                              | 14         | –          |
| Titanidae     | 3                | 4    | 0.63                          | 7.45                              | 1          | –          |
| Trachelidae   | 2                | 2    | 0.89                          | 17.68                             | –          | –          |
| Uloboridae    | 4                | 4    | 0.07                          | 8.72                              | –          | –          |
| TOTAL         | 1018             | 1359 | 0.78                          | 7.85                              | 197        | 15         |
The relationship between the BIN count for a species and the number of specimens analysed for species with a split (Fig. 6b), but the coefficient of determination was not strong ($R^2 = 0.19$). Examination of the three outlier species revealed large disparities in the abundance of their component BINs, often coupled with differing distributions as discussed below.

**Grammonota angusta.** Described in 1959 from Nova Scotia, this Nearctic species is widely distributed in Canada (Paquin et al. 2010) and has no recognized synonym (Fig. 7a). Two of its 22 BINs were common (100, 252 records), while three were of intermediate abundance, being represented by 9–11 specimens. The other 16 BINs possessed five or fewer records, including eight represented by just a single specimen. The most common BIN had the broadest distribution, occurring in all eastern provinces as well as in Manitoba, Saskatchewan and Alberta. By contrast, the remaining BINs were restricted to eastern Canada, none occurring west of Ontario. Small sample sizes made it difficult to determine the ranges of these uncommon BINs, but several were collected from distant locales (e.g. Nova Scotia & Ontario). Inspection of all sequences indicated that most substitutions were confined to third codon positions, the pattern of variation expected if the variants represent the authentic mitochondrial gene rather than a NUMT. Twelve of the 14 BINs in *G. angusta* represented by two or more specimens possessed high bootstrap values.

**Tetragnatha versicolor.** Reported from Central America to Arctic Canada, *T. versicolor* is one of the most widespread species of Tetragnathidae (Dondale et al. 2003) (Fig. 7b). Because of its morphological variability, including differences in genitalic morphology (Levi 1981), *T. versicolor* is very likely to represent a species complex.
Family | Species | BIN | Sequences | Minimum sequence divergence
---|---|---|---|---
Agelenidae | Agelenopsis pennsylvanica | BOLD:AAB5726 | 10 | 0.46
A. potteri | 106
Linyphiidae | Sisicottus montanus | BOLD:AAA8647 | 126 |
S. nesides | 29 | 0.00
S. panopeus | 57 | 0.00, 0.00
Hypselistes florens | BOLD:AAB4233 | 110 | 1.30
H. jacksoni | 34
Erigone aletris | BOLD:AAD1746 | 82 | 0.00
E. dentosa | 5
E. atra | BOLD:ACE5877 | 13 | 0.73
E. dentigera | 5
Walckenaeria arctica | BOLD:ACI8193 | 7 | 0.00
W. vigilax | 5
Lykosidae | Pardosa furcifera | BOLD:AAA5090 | 103 |
P. labradorensis | 2 | 0.00
P. medica | 19 | 0.92, 0.92
P. podhorskii | 14 | 0.46, 0.61, 0.77
P. algens | BOLD:AAA9651 | 4 |
P. glacialis | 37 | 0.83
P. sodalis | 9 | 1.24, 0.46
P. anomala | BOLD:ACE4923 | 17 | 1.08
P. sinistra | 9
P. bucklei | BOLD:ACE7902 | 24 | 0.00
P. groenlandica | 55
P. dorsalis | BOLD:ACF0659 | 163 | 1.10
P. mackenziana | 75
Alopecosa koponeni | BOLD:ACF3224 | 10 | 0.93
A. pictilis | 6
Philodromidae | Tibellus maritimus | BOLD:AAA7188 | 169 | 0.00
T. oblongus | 15
Salticidae | Hentzia mitrata | BOLD:AAC5871 | 26 | 0.00
H. palmarum | 10
Neon nelli | BOLD:AAD9221 | 21 | 0.00
N. reticulatus | 19

Table 3 Twelve pairs, two triads and one tetrad of Canadian spiders involved in BIN sharing. The minimum sequence divergence values are the lowest values between specimens assigned to the species in a particular BIN. All pairwise distances are shown for BINs which include three or four species.

Fig. 5 BIN counts for the 197 species of Canadian spiders whose members were assigned to two or more BINs.

As seven species have been assigned synonymy with it (World Spider Catalog 2014), some of the BINs detected in this study may represent named taxa. Two of the 20 BINs in T. versicolor were common (75, 125 records), while six others were represented by 10–53 specimens. The other 12 BINs were represented by five or fewer specimens. Bootstrap analysis indicated that 12 of the 14 BINs with two or more members had high support. Although some BINs represented by large numbers of specimens were restricted to either western (BOLD:AAB8024) or eastern (BOLD:AAN6690) Canada, several others had transcontinental distributions, meaning that several BINs of T. versicolor were present in most provinces.

Evarcha hoyi. Occurring across southern Canada, E. hoyi has at least one and possibly several synonyms (Marusik & Logunov 1998; World Spider Catalog 2014) (Fig. 7c). Two of its 11 BINs were common (18 records each), while the others were represented by just 1–4 specimens. One of the common BINs was restricted to British
Columbia, while the other was collected at sites across Canada. Three of the four BINs represented by several specimens were only collected in one province (two in British Columbia, one in New Brunswick), but the other was found in both Ontario and New Brunswick. Six of the seven BINs in *E. hoyi* with multiple specimens possessed strong bootstrap support.

**Geographic variation in barcode sequences**

Although nearly 28 000 barcode records were obtained, many of the species and BINs were represented by too few specimens or were collected from too few localities to permit analysis of the relationship between geographic separation and barcode sequence divergence. Analysis examined the 339 species represented by 10 or more specimens collected from sites at least 1000 km apart. Among these species, 188 did not show a significant correlation between geographical distance and sequence divergence, while the other 151 did ($P < 0.05$). The latter taxa included 103 species assigned to a single BIN and 48 placed in two or more BINs (Fig. 8). The median increase in sequence divergence in these 151 species averaged just 0.14% per 1000 km. Only eight species showed more than 1% sequence divergence per 1000 km, and each may be a species complex as their component specimens were assigned to two or more BINs.

**Species diversity**

The generation of a barcode reference library for Canadian spiders revealed 16 undescribed species as evidenced by their marked sequence and morphological divergence from known members of their genus. One of these species has now been described (*Alopecosa koponeni*, Blagoev & Dondale 2014), and descriptions are in progress for the others. As well, 52 species were newly recorded for Canada, generating first records for seven genera and one family (Table 4). Although most of these species were initially recognized through morphology, six species were only revealed when it was noted that their barcodes matched Eurasian specimens. A total of 426 new species records were added for provincial or territorial faunas (Fig. 9), in one case (Prince Edward Island) doubling the species count. In fact, the barcode-determined species occurrence records tripled the publicly available data on Canadian spiders (GBIF, accessed November 2014), illustrating the value of data releases based on DNA barcodes (Fernandez-Triana et al. 2014).

Results for the genus *Neriene* (Linyphiidae) illustrate the way in which DNA barcodes can reveal overlooked taxonomic diversity. Morphological study indicated the presence of all six species of *Neriene* reported from Canada as well as a single specimen divergent from any known taxon, a fact reinforced by its distinct barcode (Fig. 10). Specimens in five of the known species were assigned to a single BIN, but those of *N. clathrata* (Sundevall 1830) were placed in two BINs with a minimum divergence of 9.69%: one represented by six specimens from Ontario and the other by 19 from several provinces. *N. clathrata* is broadly distributed in Canada (Paquin et al. 2010), but was described from Europe so its range is Holarctic. The individuals in the BIN collected at sites across Canada (BOLD:AAA8358) showed close barcode similarity to European specimens, while those in the other BIN, BOLD:AAB7327, showed marked sequence divergence (Fig. 10). Morphological study (Fig. S2, Supporting information) indicated that males of the widely distributed BIN show close similarity to European *N. clathrata* (Nentwig et al. 2014), while those of the other BIN resemble the male of *Neriene waldea* Chamberlin & Ivie 1943, a species synonymized with *N. clathrata* nearly 50 years ago (van Helsdingen 1969).

**Automated taxonomic placements of BINs**

Although some specimens in this study could not be assigned to a known species, each of the 1359 BINs was
readily assigned to a family based on the morphology of its members. This allowed a test of the correspondence with the assignments made when each BIN was placed in the same family as its NN taxon on BOLD. The results of this analysis indicated that all BINs with <10% sequence divergence from their NN were assigned to the

Fig. 7 Condensed neighbour-joining trees for the three Canadian spiders with exceptional BIN diversity: (a) Grammonota augusta (22 BINs, 411 individuals), (b) Tetragnatha versicolor (20 BINs, 338 individuals) and c) Evarcha hoyi (11 BINs, 54 individuals).
correct family (Table 5). Most (94%) assignments were also correct for NN distances from 10 to 15%, but BINs with higher NN distances were often misassigned.

**Discussion**

The present study has assembled DNA barcode sequences for 1018 species of spiders, two-thirds of the Canadian fauna. Because 98% of these species possessed a diagnostic array of sequences, the present DNA barcode reference library is highly effective in identifying Canadian spiders. In those few cases where it fails to deliver a species-level identification, it assigns a specimen to a few closely related species. The development of this barcode library has codified the expertise of the leading taxonomists working on Canadian spiders in a format that will allow more researchers to explore the diversity of this group.

This study revealed that nearest-neighbour distances were $10 \times$ higher for Canadian spiders than intraspecific divergences (7.85% vs. 0.78%). These measures of variation presume that the current taxonomic system is valid and that all specimens have been correctly identified, requirements impossible to fully satisfy without further study. The few species lacking barcode divergence need to be investigated to validate their reproductive isola-
Table 4 Sixty-eight species of spiders new for Canada, including 16 undescribed species and 52 previously known from another nation

| Family        | Species                        | No. of species | Identifier | Occurrence in Canada | General distribution |
|---------------|--------------------------------|----------------|------------|----------------------|----------------------|
| Agelenidae    | *Agelenopsis naevia*           | 2♂, 4♀        | 1          | ON                   | USA                  |
| Anapidae      | *Comaroma mendocino*           | 1♀             | 2          | BC                   | USA                  |
| Antrodiaetidae| *Antrodiaetus cerberus*        | 4♂             | 2          | BC                   | USA                  |
| Araneidae     | *Acanthepeira venusta*        | 5♂, 15♀       | 1          | ON                   | USA, Cuba, Hispaniola |
|               | *Eustala emertonii*           | 3♂, 3♀        | 1          | SK, ON, QC, NS       | USA, Mexico          |
|               | *Eustala rosae*               | 19♀, 41♀      | 1          | ON, NB, PE           | USA, Mexico          |
|               | *Ges heptagon*                | 3♂, 3♀        | 1          | USA to Argentina, South Pacific Is., Australia |
| Cybaeidae     | *Cybaeus* sp. 1RB†            | 1♂, 1♀        | 2          | BC                   | –                    |
| Dictynidae    | *Dictyna bellans*             | 3♀             | 1          | ON                   | USA, Mexico          |
|               | *Dictyna tetricus*            | 1♀             | 1          | BC                   | Europe               |
| Eutichuridae  | *Strotarchus piscatorius*     | 7♂             | 5          | ON, NS               | USA, Mexico          |
| Gnaphosidae   | *Gnaphosa fontinalis*         | 1♀             | 1          | ON                   | North, Central America |
|               | *Nodocion mateonus*           | 1♂, 1♀        | 1          | BC, SK               | USA                  |
|               | *Zelotes pseustes*            | 1♂, 3♀        | 1          | ON                   | USA, Mexico          |
|               | *Zelotes rainier*             | 1♂, 1♀        | 2          | BC                   | –                    |
| Linyphiidae   | *Agyneta cauta*               | 1♂             | 1          | AB                   | Paleartic            |
|               | *Agyneta dolomii* cf.†        | 1♂, 1♀        | 1          | BC                   | –                    |
|               | *Agyneta decorata*            | 1♂, 5♀        | 1          | BC                   | USA                  |
|               | *Agyneta filiformis* cf.†     | 1♂             | 1          | BC                   | –                    |
|               | *Agyneta fibuscrocus*         | 1♂, 7♀        | 1          | BC                   | USA                  |
|               | *Agyneta manni*               | 1♂, 24♀       | 2          | BC                   | USA                  |
|               | *Agyneta ripariensis*         | 6♂             | 3          | YT                   | Russia               |
|               | *Agyneta sp. 1GAB†*           | 1♀             | 1          | AB                   | –                    |
|               | *Agyneta sp. 2GAB†*           | 19             | 1          | AB                   | –                    |
|               | *Agyneta sp. 3RB†*            | 1♂, 1♀        | 1          | BC                   | –                    |
|               | *Agyneta uta*                 | 2♂, 5♀        | 1          | BC                   | USA                  |
|               | *Asiceratinops kolymensis§*   | 1♂, 19, 2♀     | 4          | MB, ON               | Russia               |
|               | *Centromerus sp. 2GAB†*       | 6♂, 19♀       | 1          | BC                   | –                    |
|               | *Centromerus sp. 3GAB†*       | 1♂             | 1          | BC                   | –                    |
|               | *Eulaira obscura*             | 7♂, 139, 2♀   | 1          | BC, AB               | USA                  |
|               | *Gibothorax tchemoviš*        | 8♀             | 3          | YT                   | Russia               |
|               | *Improphantes sp. 1GAB†*      | 1♂, 119, 7♀   | 1          | BC, AB               | –                    |
|               | *Improphantes sp. 1RB†*       | 7♂, 38♀       | 1          | BC, AB               | –                    |
|               | *Linyphantes pacificus*       | 3♂, 119♀      | 1          | BC                   | USA                  |
|               | *Micrhiphanes* sp. 1GAB†§‌    | 1♂, 49, 4♀    | 1          | BC, AB, MB, QC, NF   | –                    |
|               | *Oreonetides sp. 1GAB†*       | 2♂, 149, 2♀   | 1          | BC, AB               | –                    |
|               | *Oreonetides sp. 2GAB†*       | 5♂, 24♀       | 1          | BC, AB               | –                    |
|               | *Scotinotylus evansi*         | 1♀             | 5          | MB                   | Greenland, Paleartic |
|               | *Scotinotylus kenus* cf.†     | 2♂, 12♀       | 1          | BC, AB               | –                    |
|               | *Semiijica alctica*           | 2♂, 5♀        | 3          | YT                   | Sweden, Finland, Russia |
|               | *Silomotopus reussi*          | 1♀             | 5          | BC                   | Paleartic            |
|               | *Tapnoctea affinis* cf.†      | 3♂, 3♀        | 1          | BC, AB               | –                    |
|               | *Tapnoctea hortensis*         | 1♂, 1♀        | 1          | ON                   | USA                  |
|               | *Tenuiphanites mengei*        | 1♀             | 5          | ON                   | Paleartic            |
|               | *Walckenaeria occidentalis*   | 1♂, 2♀        | 1          | BC                   | USA                  |
| Lycosidae     | *Alopecosa koponenii†*        | 4♂, 2♀, 4♀    | 1          | MB                   | –                    |
|               | *Alopecosa mutabilis*         | 3♀             | 5          | NU                   | Russia, Alaska       |
|               | *Alopecosa muntzii*           | 1♂, 1♀        | 1          | ON                   | USA                  |
tion by testing for sequence divergence at nuclear loci, ideally at localities where they are sympatric. If no differences are detected, some of these taxa may reflect oversplitting; if differences are present, they may indicate cases of incomplete lineage sorting or introgressive hybridization. However, because cases of sequence sharing are so uncommon, representing <1% of all taxa, their resolution will have little impact on the overall metrics of DNA barcode variation. There is a much greater need to understand the factors responsible for the many cases of deep intraspecific sequence divergence which led 16.5% of the species to be placed into two or more BINs. Because of the strong correspondence between BINs and traditionally recognized species (Ratnasingham & Hebert 2013; Huemer et al. 2014; Pentinsaari et al. 2014; Zahiri et al. 2014), many of these BIN splits may represent species overlooked by the current taxonomic system, but other explanations are

| Family     | Species                                   | No. of species | Identifier | Occurrence in Canada | General distribution |
|------------|-------------------------------------------|----------------|------------|----------------------|---------------------|
| Mimetidae  | Mimetus haynesi                           | 1♂             | 1          | ON                   | USA                 |
| Oxyopidae  | Oxyopes salticus                         | 7♂, 6♀        | 1          | ON                   | USA to Brazil       |
| Pholcidae  | Pholcus manueli                          | 5♂, 8♂, 1 I    | 1          | ON                   | Russia, Turkmenistan, China, Korea, Japan, USA |
| Salticidae | Pholcus opilionoides                      | 5♂, 3♂, 6 I    | 1          | BC, ON               | Europe to Azerbaijan |
|            | Dendryphantes czekanowskii               | 1♀             | 1          | YT                   | Russia              |
|            | Habronattus agilis                       | 1♂, 5♂, 3 I    | 1          | ON                   | USA                 |
|            | Myrmarachne formicaria                   | 2♂             | 1          | ON                   | Palearctic, USA (introduced) |
|            | Sitticus conceler                        | 3♂, 19, 1 I    | 1          | ON                   | USA, Mexico         |
|            | Sitticus pubescens                       | 2♂             | 1          | ON                   | Europe, Russia, USA |
|            | Terralonus mylothuris                    | 1 I            | 5          | AB                   | USA                 |
| Theridiidae| Dipoena procalis                         | 19             | 2          | BC                   | USA                 |
|            | Stratoda woshoni                         | 29             | 2          | BC                   | USA, Mexico         |
|            | Theridion mystaceum                      | 19             | 2          | BC                   | Palearctic          |
| Thomisidae | Synema parzialum                         | 2♂, 29         | 1          | ON                   | USA, Mexico         |
| Uloboridae | Uloborus diversus                        | 19, 1 I        | 1          | ON                   | USA, Mexico         |

Species: †undescribed species (n = 16), ‡new family record (n = 1) and §new genus record (n = 7).
Identifiers: G Blagoev (1), R Bennett, C & D Copley (2), Y Marusik (3), D Buckle (4) and DNA barcoding (5).

Fig. 9 Number of species of spiders with barcode records, total known species and new species reported for each Canadian province or territory.
Fig. 10 Neighbour-joining tree for six named and one unnamed species of *Neriene* from Canada. As *N. clathrata* was assigned to two BINs, sequences from 23 European specimens were included for comparison. *Frontinella communis* is a closely related outgroup.

possible [e.g. biased variation stemming from infection with maternally transmitted parasites (Kvie et al. 2013)]. These cases have the potential to substantially shift key barcode metrics (mean intraspecific variation, mean NN distance) and overall species diversity. The incidence of BIN splits (19.3%) detected in this study is considerably
higher than that reported in prior work on other groups of arthropods (e.g. noctuoid Lepidoptera – 9.1% (Zahiri et al. 2014); Coleoptera – 5.0–5.7% (Hendrich et al. 2014; Pentinsaari et al. 2014)), but reinforces earlier evidence for the prevalence of deep intraspecific divergences in spiders (Barrett & Hebert 2005; Blagoev et al. 2009, 2013). In fact, maximum sequence divergence exceeded 4% in 95 of the 876 species represented by multiple individuals and more than 2% in 322 of these taxa. The present study revealed three species with exceptional BIN diversity. For example, specimens of Tetragnatha versicolor possessed up to 9.69% divergence and were assigned to 20 BINs which showed evidence of differing geographic distributions. Because three to five lineages co-occur at many sites across Canada, tests for their reproductive isolation in sympatry should be possible. Although no other studies have aimed to quantify the incidence of terrestrial arthropods with high BIN diversity, other cases have been detected. For example, one species of Canadian noctuoid among 1531 species was assigned to 16 BINs (Zahiri et al. 2014).

The present study revealed 16 undescribed species with clear barcode and morphological divergence from any known species, but they may only represent the tip of a taxonomic iceberg. Future studies need to examine the many species of Canadian spiders assigned to more than one BIN to ascertain if most of these cases represent a cryptic species complex. These follow-up studies could involve more detailed morphological investigation, and perhaps the inclusion of one or more nuclear markers to test for reproductive isolation. In addition, these analyses should exclude alternative explanations for the sequence divergence patterns, such as geographic structure, biased variation induced by maternally transmitted parasites (Kvie et al. 2013), heteroplasmy (Frey & Frey 2004) or co-amplification of pseudogenes (Song et al. 2008). If these follow-up studies similarly corroborate the barcode results (e.g. Neriene clathrata, Fig. S2, Supporting information), this would imply that 30% of the fauna has been overlooked. Moreover, because half of the species (536 of 1018) examined in this study were represented by fewer than 10 specimens, it is likely that many barcode splits await detection. Expanded sample sizes might well reveal that the current species count for Canadian spiders is a 50% underestimate.

At least 40 000 species of spiders are thought to await description, a task whose completion has been estimated to cost $500M and to require another half century (Platnick & Raven 2012). By suggesting that the current taxonomic system overlooks many closely related species, the present study implies that the actual number of spiders awaiting description is larger, potentially raising costs and time. However, by showing the close correspondence between BINs and species, this study has revealed a pathway that can speed the registration of new species and lower the cost of such work. The Canadian spider fauna is better known and less diverse than many other regions, but expect this pathway to be universal. There is no evidence to suggest that the results would be markedly different in more diverse areas, or regions with fragmented and complex habitats, but this remains to be tested. The present results indicate that nearly 80% of the sequence clusters recognized through BIN analysis coincide with recognized species. Moreover, the present study shows that new BINs can gain automated taxonomic placements, as evidenced by the accuracy of family assignments in cases where NN divergences were less than 10%. As the barcode reference library gains parameterization, fewer species will exceed this threshold, and more newly encountered BINs will also be deeply embedded in clades comprised of species belonging to a particular subfamily or a genus, allowing the automation of more precise taxonomic placements. Given the accelerating pace of species loss, the need for speed in mapping species distributions and registering species diversity is obvious. The synoptic approach to species diversity provided by BINs can hasten biodiversity evaluations at precisely the time when this capacity is critically needed.

### Acknowledgements

Financial support was provided by the Ontario Ministry of Research and Innovation and by the government of Canada through Genome Canada and the Ontario Genomics Institute in support of the International Barcode of Life project. We are extremely grateful to colleagues who provided specimens for analysis (Charles Dondale, Chris Buddle, Claudia Copley, Darren Copley, John Sloan, Robb Bennett, Tom Mason, Vadim Junea) and to those who shared their taxonomic expertise (Andrei Tanasevitch, Charles Dondale, Christo Deltchev, Donald Buckle, Nadine Duppérel, Robb Bennett, Theo Blick, Yuri Marusiak, Jozef Slowik). We particularly thank Robb Bennett as well as Claudia and Darren Copley whose survey and curatorial work on Canadian spiders are unrivalled. We are very grateful to John Waithaka and other staff at Parks Canada for supporting collection programs in national parks across Canada. We also thank colleagues at the Biodiversity Institute of Ontario (particularly Jayme Sones, Renee Labbee, Crystal Sobel, Kate Perez, and Mon-

| Barcode divergence value (K2P) | n   | Correct assignments (%) |
|-------------------------------|-----|-------------------------|
| 0.00–4.99%                    | 264 | 100                     |
| 5.00–9.99%                    | 443 | 100                     |
| 10.0–14.99%                   | 280 | 94                      |
| 15.0–24.99%                   | 31  | 39                      |

| Barcode divergence value (K2P) | n   | Correct assignments (%) |
|-------------------------------|-----|-------------------------|
| 0.00–4.99%                    | 264 | 100                     |
| 5.00–9.99%                    | 443 | 100                     |
| 10.0–14.99%                   | 280 | 94                      |
| 15.0–24.99%                   | 31  | 39                      |
ica Young) who played a critical role in making the collections, in sorting samples and in carrying out sequence and data analyses. Finally, we wish to thank the three anonymous reviewers whose comments greatly improved our manuscript.

**Ethics statement**

All animal work was conducted according to relevant national and international guidelines. No specific permissions are required to work with invertebrates in Canada. For field collection, no specific permissions are required for the collection of spiders from public areas. Specimens collected within Canadian National Parks were performed so under permits NAP-2008-1636 and PC-2012-11074 granted by Parks Canada. None of the field collections involved endangered or protected species.

**References**

Astrin JJ, Huber BA, Misof B, Klutsch CFC (2006) Molecular taxonomy in pholcid spiders (Pholcidae, Araneae): evaluation of species identification methods using COI and 16s RNA. Zoologica Scripta, 35, 441–457.

Barrett RDH, Hebert PDN (2005) Identifying spiders through DNA barcoding. Canadian Journal of Zoology, 83, 481–491.

Bennett R (1999) Canadian spider diversity and systematics. Newsletter of the Biological Survey of Canada, 38, 17–27.

Blagoev GA, Dondale CD (2014) A new species of Alopecosa (Araneae: Lycosidae) from Canada: a morphological description supported by DNA barcoding of 19 congers. Zoootaxa, 3894, 152–160.

Blagoev G, Hebert PDN, Adamowicz S, Robinson E (2009) Prospects for using DNA barcoding to identify spiders in species-rich genera. ZooKeys, 16, 27–46.

Blagoev GA, Nikolova NI, Sobel CN, Hebert PDN, Adamowicz SJ (2013) Spiders (Araneae) of Churchill, Manitoba: DNA barcodes and morphology reveal high species diversity and new Canadian records. BMC Ecology, 13, 44.

Candek K, Kuntner M (2014) DNA barcoding gap: reliable species identification over morphological and geographical scales. Molecular Ecology Resources, 15, 268–277.

Dondale CD (1979) Araneae. In: Canada and Its Insect Fauna (ed. Danks HV), 108, pp 247–250. Memoirs of the Entomological Society of Canada, Cambridge University Press, Cambridge, 573 pp.

Dondale CD, Redner JH, Paquin P, Levi HW (2003) The insects and arachnids of Canada. Part 23. The orb-weaving spiders of Canada and Alaska (Araneae: Uloboridae, Tetragnathidae, Araneidae, Theridiosomatidae), p. 371. NRC Research Press, Ottawa.

Duperré N (2013) Taxonomic revision of the spider genera Agyneta and Teneuseeum (Araneae, Linyphiidae) of North America north of Mexico with a study of the embolic division within Micronetinae sensu Saaristo & Tanasewitch 1996. ZooKeys, 3674, 1–189.

Fernandez-Triana J, Fenev L, Ratnasingham S et al. (2014) Streamlining the use of BOLD specimen data to record species distributions: a case study with ten Nearctic species of Microgastrinae (Hymenoptera: Braconidae). Biodiversity Data Journal, 2, e1153.

Foelix RF (2011) Biology of spiders, 3rd edn. Oxford University Press, New York, 419 pp.

Folmer O, Black M, Hoeh W, Lutz R, 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.

Freitas GCDC, Brescovit AD, Vasconcelos SD (2013) Spider diversity on the oceanic island of Fernando De Noronha, Brazil, and implications for species conservation. Journal of Insect Science, 13, 148.

Frey JE, Frey B (2004) Origin of intra-individual variation in PCR-amplified mitochondrial cytochrome oxidase I of Thriss tubaci (Thysanoptera: Thripidae): Mitochondrial heteroplasmy or nuclear integration? Hereditas, 140, 92–98.

Fujiwara T, Barraclough TJ (2013) Delimiting species using single-locus data and the generalized mixed yule coalescent approach: a revised method and evaluation on simulated data sets. Systematic Biology, 62, 707–724.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. Proceedings of the National Academy of Sciences of the United States of America, 103, 9684–9671.

Hajibabaei M, Shokralla S, Zhou X, Singer GA, Baird DJ (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS ONE, 6, e17497.

Hausmann A, Haszprunar G, Hebert PDN (2011) DNA barcoding the geometrid fauna of Bavaria (Lepidoptera): successes, surprises and questions. PLoS ONE, 6, e17134.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences, 270, 313–321.

Hebert PDN, Benton EH, Burns J, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly, Astraptes fulgerator. Proceedings of the National Academy of Sciences of the United States of America, 101, 14812–14817.

Hebert PDN, deWaard JR, Landry JF (2009) DNA barcodes for 1/1000 of the animal kingdom. Biology Letters, 5, 359–362.

Hebert PDN, de Waard JR, Zakharov EV et al. (2013) A DNA ‘Barcode Blitz’: Rapid digitization and sequencing of a natural history collection. PLoS ONE, 8, e60535.

van Helsing P, Moriniere J, Haszprunar G et al. (2014) A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3500 identified species to BOLD. Molecular Ecology Resources, 15, 795–818.

Huemer P, Mutanen M, Sefc KM, Hebert PDN (2014) Testing DNA Barcoding Performance in 1000 Species of European Lepidoptera: Large geographical Distances Have Small Genetic Impacts. PLoS ONE, 9, e115774.

Ivanova NV, deWaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. Molecular Ecology Notes, 6, 998–1002.

Ji Y, Ashton L, Pedley SM et al. (2013) Reliable, verifiable, and efficient monitoring of biodiversity via metabarcoding. Ecology Letters, 16, 1245–1257.

Kiev KS, Hogner S, Aarvik L, Lifjeld JT, Johnsen A (2013) Deep sympatric mtDNA divergence in the autumnal moth (Epirrita autumnata). Ecology and Evolution, 3, 126–144.

Levi HW (1981) The American orb-weaver genera Dolichognatha and Tetragnatha north of Mexico (Araneae: Araneidae, Tetragnathinae). Bulletin of the Museum of Comparative Zoology at Harvard College, 149, 271–318.

Mantel NA (1967) The detection of disease clustering and a generalized regression approach. Cancer Research, 27, 209–220.

Marusik YM, Logunov DV (1998) Taxonomic notes on the species complex (Araneae Salticidae). Arthropoda Selecta, 6, 95–104.

Miller JA, Belgers JDM, Beentjes KK, Zwakhals K, van Helsing P (2013b) Spider hosts (Arachnida, Araneae) and wasp parasitoids (Insecta, Hymenoptera, Ichneumonidae, Epiplaiini) matched using DNA barcodes. Biodiversity Data Journal, 1, e992.
G.A.B., J.R.D. and P.D.N.H conceived and designed the experiments. G.A.B., S.L.D. and J.R.D. performed the experiments. G.A.B., J.R.D., S.R., S.L.D., J.R. and A.C.T. analysed the data. G.A.B., J.R.D., S.R., S.L.D., L.L., J.R., A.C.T. and P.D.N.H. contributed reagents/materials/analysis tools. G.A.B., J.R.D. and P.D.N.H wrote the study. All authors read and approved the final manuscript.

Data accessibility

All collection and sequence data are available on BOLD (www.boldsystems.org) (Ratnasingham & Hebert 2007) in the public data set DS-SOC2014 (doi:10.5883/DS-SOC2014). The distance matrices for Fig. 8, the COI sequence alignment and the BOLD neighbour-joining tree (as both PDF and Newick files) have all been uploaded to DRYAD (doi:10.5061/dryad.21364). The sequence and primer data are also available on GenBank (see Table S1, Supporting information for accession numbers).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 BOLD neighbour-joining tree (K2P) for all 27,269 COI sequences >500 bp from spiders collected in Canada.

Fig. S2 Retrolateral and prolateral views of left pedipalps for males of Nerine cleftbura showing diagnostic differences between a) BIN BOLD:AAA8358 and b) BIN BOLD:AAAB7327.

Table S1 Summary data for the 27,269 specimens of Canadian spiders analysed, including BOLD and GenBank accessions, BIN assignments, and locality details.

Table S2 197 species of Canadian spiders that were assigned to two or more BINS.