

# Barcoding the Collembola of Churchill: a molecular taxonomic reassessment of species diversity in a sub-Arctic area

DAVID PORCO,\* DARIUSZ SKARŻYŃSKI,† THIBAUD DECAËNS,\* PAUL D. N. HEBERT‡ and LOUIS DEHARVENG§

\*EA 1293 ECODIV, Université de Rouen, UFR Sciences et Techniques, SFR SCALE, Bâtiment IRESE A, Mont Saint Aignan Cedex 76821, France, †Department of Evolutionary Biology and Ecology, Wrocław University, Przybyszewskiego 63/77, Wrocław 51-148, Poland, ‡Biodiversity Institute of Ontario, University of Guelph, 50 Stone Road East, Guelph, ON N1G 2W1, Canada, §Muséum National d'Histoire Naturelle, UMR7205, «Origine, Structure et Evolution de la Biodiversité», 45 rue Buffon, CP50, Paris 75005, France

## Abstract

Although their functional importance in ecosystems is increasingly recognized, soil-dwelling micro-arthropods are usually poorly known in comparison with their above-ground counterparts. Collembola constitute a significant and species-rich component of the soil biodiversity, but it remains a woefully understudied group because of the taxonomic impediment. The ever-increasing use of molecular taxonomic tools, such as DNA barcoding, provides a possible solution. Here, we test the use of this approach through a diversity survey of Collembola from the vicinity of Churchill, Manitoba, Canada, and compare the results with previous surveys in the same area and in other sub-Arctic regions. The systematic barcoding campaign at Churchill revealed a diverse collembolan fauna consisting of 97 species-level MOTUs in six types of habitats. If all these MOTUs are confirmed as species, this richness would be far higher than prior records for Arctic Canada and could lead to reconsider the actual diversity of the group in Arctic environments.

**Keywords:** Arctic, Collembola, cryptic diversity, diversity, DNA barcoding, species proxy

Received 24 June 2013; revision received 9 September 2013; accepted 13 September 2013

## Introduction

The class Collembola is one of the most abundant groups of decomposers in soil. With a global diversity of nearly 8000 described species, it occurs in a broad range of habitats, including the most extreme environments (Hopkin 1997). Even in well-explored areas, such as Western Europe, species new to science are still being discovered (Deharveng 2004), and many more are expected in tropical areas usually accepted as a rich and underinvestigated source of biodiversity in many groups (May 2010). The Neotropical zone currently includes 1557 described species of Collembola (Mari Mutt & Bellinger 1990, 1996). By contrast, Arctic environments have generally been found less rich in species: only 420 species of Collembola are known from the entire Arctic (Babenko & Fjellberg 2006). But the species richness of the different biomes are currently underestimated (Deharveng 2004).

Applying the ratio found by Mora *et al.* (2011) between described and unknown species for the whole animal kingdom, we estimate that the global diversity of Collembola could reach about 65 000 species. The key reason for the current underestimation of collembolan diversity, as in many other soil invertebrates, is the taxonomic impediment (Decaëns 2010), that is, difficult taxonomy combined with a paucity of specialists. DNA barcoding, a molecular taxonomic tool, using a portion of the 5' end of the mitochondrial gene COI as a species tag (Hebert *et al.* 2003) provides a possible solution. The systematic use of this tool for the development of extensive reference libraries in Collembola has shown that even in temperate-zone environments and for common species, considered well known, a significant part of the diversity has been overlooked (Cicconardi *et al.* 2010; Porco *et al.* 2012a,b). These results suggest that the global diversity of Collembola is higher than expected, with many species new to science awaiting description in regions that were considered to be well explored, such as Western Europe. As a consequence, the current species count for

Correspondence: David Porco, Fax: +33 2 35 14 66 55; E-mail: david.porco.gm@gmail.com

Arctic regions, which have been less sampled, could likely be a great underestimate as well. Moreover, in the context of global warming, the accurate estimation of the species communities present in Arctic ecosystems is particularly critical as it will impact the cryophilic native fauna, favour introduced species, possibly affecting both ecosystem functioning and services delivery.

This study presents information on the diversity of collembolan species present near Churchill (Manitoba, Canada). The sampling activity and the sequencing of the specimens were carried out in the context of a large-scale DNA barcoding study of polar life (PolarBOL – <http://www.polarbarcoding.org>). This survey of the collembolan fauna combined morphological analysis with DNA barcodes to provide a comprehensive appraisal of the local diversity of this group in an Arctic environment.

## Materials and methods

### *Study site*

The Churchill area is located in northern Manitoba, in the Canadian sub-Arctic region, on the shores of Hudson

Bay near the estuary of the Churchill River. This region sits on the interface between boreal forest and tundra. Habitats with a maritime influence such as coastal tundra, bluff and beach are also present.

### *Sampling*

Two sampling trips focused on the springtail fauna (July 2008 and July 2009), but extra samples were collected during previous generalist expeditions (September 2000, August 2006, June 2007) (Table S1). The collembolan fauna from soil, moss, lichen and rotten wood was extracted using Berlese funnels. In addition, sweep net, hand collecting, pitfall traps and bush beating were used to complete the sampling for epiedaphic species. Six types of habitats (coastal tundra, tundra, beach, bluff, forest and fen) were investigated in the Churchill area and its vicinity, for a total of 48 sampling localities (Table S1, Fig. 1). From 3 to 20 replicates were sampled for each habitat type. Except for the fen, the sampling effort was proportional to the representation of those habitats in the surrounding landscape (Table 1). All samples were preserved in 95% ethanol. The specimens were sorted

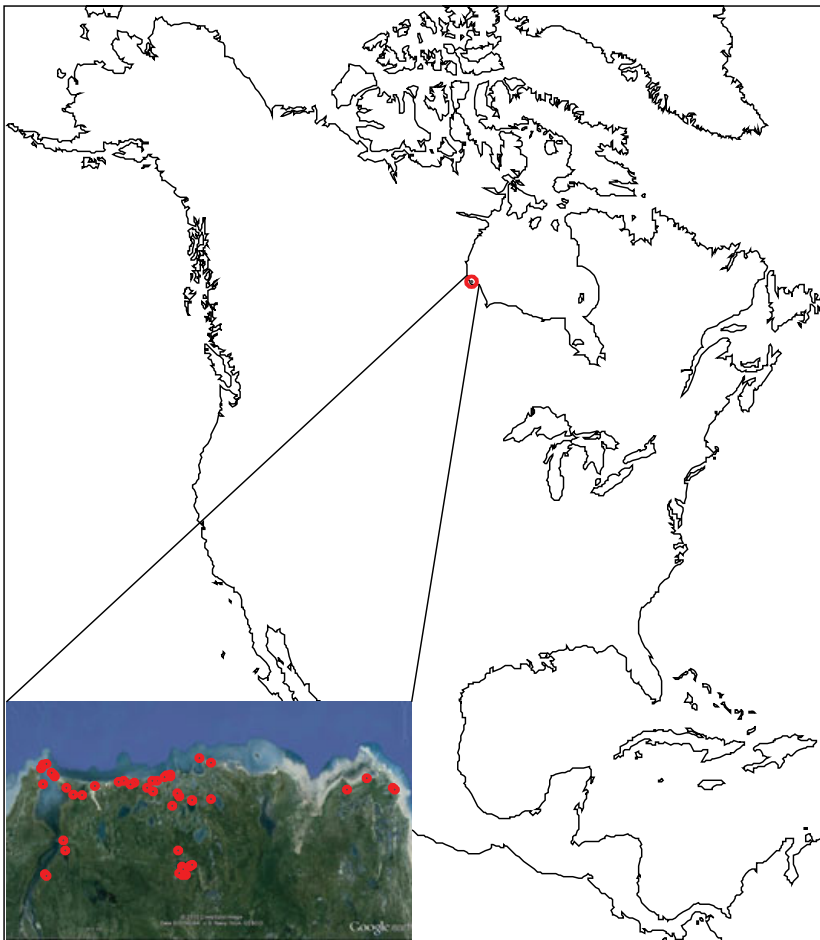


Fig. 1 Map of sampling sites.

**Table 1** Number of replicates and specimens collected per habitat

Habitats	No. of replicates	N
Forest	3	125
Fen	4	26
Bluff	5	123
Beach	7	109
Coastal tundra	14	260
Tundra	20	423

into morphospecies, and several individuals of each were sequenced from each sampling site. A total of 1066 specimens were sequenced for the 48 sampling sites (Table S1). Identified specimens were deposited in the Museum of Natural History in Paris. Except for unique specimens, duplicates of the vouchers were deposited at the Biodiversity Institute of Ontario.

### Molecular analysis

DNA was extracted from entire specimens in 30  $\mu$ L of lysis buffer ([http://www.ccdb.ca/docs/CCDB\\_DNA\\_Extraction.pdf](http://www.ccdb.ca/docs/CCDB_DNA_Extraction.pdf)), and proteinase K incubated at 56°C overnight. DNA extraction followed a standard automated protocol using 96-well glass fibre plates (Ivanova *et al.* 2006). Specimens were recovered after DNA extraction using a specially designed work flow allowing their morphological examination (Porco *et al.* 2010a). Species were identified using standard identification keys (mainly Potapov 2001; Fjellberg 1980; Christiansen & Bellinger 1998; Bretfeld 1999). The 5' region of COI used as a standard DNA barcode was amplified using M13-tailed primers LCO1490 and HCO2198 (Folmer *et al.* 1994). Samples that failed to generate an amplicon were subsequently amplified with a pair of internal primers combined with full length ones (LepF1-MLepR1 and MLepF1-LepR1) (Hajibabaei *et al.* 2006). The standard PCR reaction protocol of the Canadian Center for DNA Barcoding was used for amplifications (<http://www.dnabarcodes2011.org/conference/preconference/CCDB-Amplification-animals.pdf>), and products were checked on a 2% E-gel 96Agarose (Invitrogen). Unpurified PCR amplicons were sequenced in both directions using M13-tailed primers (Hajibabaei *et al.* 2005), with products subsequently purified using Agencourt CleanSEQ protocol and processed using BigDye version 3.1 on an ABI 3730 DNA Analyzer (Applied Biosystems). Sequences were assembled and edited with Sequencher 4.5 (GeneCode Corporation, Ann Arbor, MI, USA). The alignment was obtained using BIOEDIT version 7.0.5.3 (Hall 1999). Sequences are publicly available on BOLD via the following DOI

(<http://dx.doi.org/10.5883/DATASET-CHU-COL1>) and on GenBank (GU657078–GU6573 35, HM390591–HM390627, HM390653–HM390662, HM 424138–HM424148, HM431644–HM431645, HM893770–HM893811, HQ992117, JF884220–JF884222, JF884258, JN 269617–JN269618, JN306349–JN306356, JX261789, JX26 1792, JX261796, JX261798, JX261801, JX261804, JX261808, JX261809, JX261814, JX261819, JX261833, JX261837, JX 261843, JX261844, JX261848, JX261857, JX261876, JX26 1884, JX261886, KF641929–KF642599).

### Data analyses

*Molecular distance analysis and MOTUs delineation.* Distance analyses were performed with MEGA5 (Tamura *et al.* 2011), utilizing a neighbour-joining (Saitou & Nei 1987) algorithm with the Kimura-2 parameter model (Kimura 1980) to estimate genetic distances. The robustness of nodes was evaluated through bootstrap re-analysis of 1000 pseudoreplicates. The tree was replotted using the online utility iTOL (Letunic & Bork 2007). MOTUs were defined with 'mothur' using Hcluster command with the option 'Furthest neighbour' (Schloss *et al.* 2009).

*Rarefaction curves and diversity estimators.* The data were analysed using rarefaction procedures that are specifically designed to avoid the potential bias generated by uneven sampling. Rarefaction curves for specific diversity were generated with EcoSim 7.71 (Gotelli & Entsminger 2006) with a 95% confidence level and plotted with R 2.15.0 (R Development Core Team 2012) using the package 'Plotrix' (Lemon 2006). Rarefaction curves were calculated separately for MOTUs and species accumulation as a function of sampling effort (number of collected individuals) in the four orders of Collembola, and for MOTUs accumulation in the six different types of habitats sampled (coastal tundra, tundra, beach, bluff, forest and fen). The observed cumulated richness was calculated as the total number of MOTUs and species observed in a given habitat. Additionally, theoretical species richness was estimated globally and for each habitat using the Chao1 and ACE diversity estimators using with the 'Vegan' package for R 2.15.0 (Oksanen *et al.* 2012).

Beta diversity between habitats was assessed using the Sørensen's index of dissimilarity:  $\beta_{BC} = (b + c) / (2a + b + c)$ , where a is the number of MOTUs shared between two sites B and C, and b and c are the numbers of unique MOTUs (not shared) for sites B and C.

Differences in assemblage composition between habitats were highlighted through a correspondence analysis (Benzécri 1973; Greenacre 1984) performed with the package 'ade4' (Chessel & Dufour 2004) for the R 2.15.0

environment. The table used for this analysis had 6 lines (habitats) and 97 columns (MOTUs).

## Results

Morphological analysis revealed the presence of 45 species: four represented by a single individual (singletons), and the 41 species represented by multiple individuals were used as a reference for the empirical definition of a MOTU delineation threshold (the rest of the dataset was mostly identified to the genus level and classified into 43 morphospecies). The calculation of the maximum intraspecific genetic divergence within each of these species revealed a clear gap between 13.45% and 20% (Fig 2). Thirty-two of the morphologically identified species exhibited maximum intraspecific distances under 10%, only two species (*P. binoculatus* and *P. ekmani*) possessed higher values (11.8% and 13.45%, respectively). Individuals of seven other species showed higher divergences (from 20.0% up to 30.0%), but this corresponded to separation between discrete genetic lineages. The mean genetic divergence between these seven lineages was 22.2% (range 15.9–27.7%), while the mean intralinear distance was 1.96% (range 0–5.58%). The genetic distances between and within these lineages are similar to values in the 38 other nominal species, which each corresponded to a different MOTU: the mean interspecific distance was 27.3% (range 16.7–42.5%) and the mean intraspecific distance was 1.3% (range 0–6.6%). Six species were only encountered in a single population, so the possible existence of other intraspecific divergent lineages remaining to be sampled in the Churchill area cannot be ruled out. Thus, these lineages appear as divergent as nominal species are from each other, and so can be

considered as specific-level MOTUs. This first step suggested that a conservative 14% threshold could be applied to delineate species-level MOTUs in the rest of the data set.

The second step of the data analysis compared results obtained from the morphologically identified species with the MOTU-based delineation for the whole data set. The estimates for the number of MOTUs obtained with different threshold values showed an initial plateau around 10% and a second at 14% value with a steep decline after 19% (Fig. 3). These plateaux, representing the insensitivity to changes in cut-off value, were already suggested as a sign that most MOTUs delineated could be valid biological species (Plaisance *et al.* 2011). Moreover, examination of the specimens composing MOTUs under different thresholds showed the agglomeration of morphologically identified MOTUs at higher threshold values. Conversely, when values lower than 14% were applied, it led to the partitioning of these MOTUs. This confirmed that a minimum threshold value of 14% could be applied to the whole data set to delineate species-level MOTUs.

When a 14% divergence threshold was adopted, a total of 97 MOTUs were retrieved with 18 singletons (Fig. 4). The mean divergence within MOTUs was 1.4% (range 0–13.9% – Table S2, Fig. S1) and between MOTUs was 28.7% (range 14.5–44.4% – Table S2, Fig. S1). Levels of variation within MOTUs were similar among the three best sampled orders (Symphypleona 0.9%; Poduromorpha 1.5%; Entomobryomorpha 1.5% – Fig. S2) with species of *Micranurida* exhibiting the highest mean value (13.09% – Table S2).

Concerning diversity assessment, a plot of the rarefaction curves, showed that an asymptote was almost

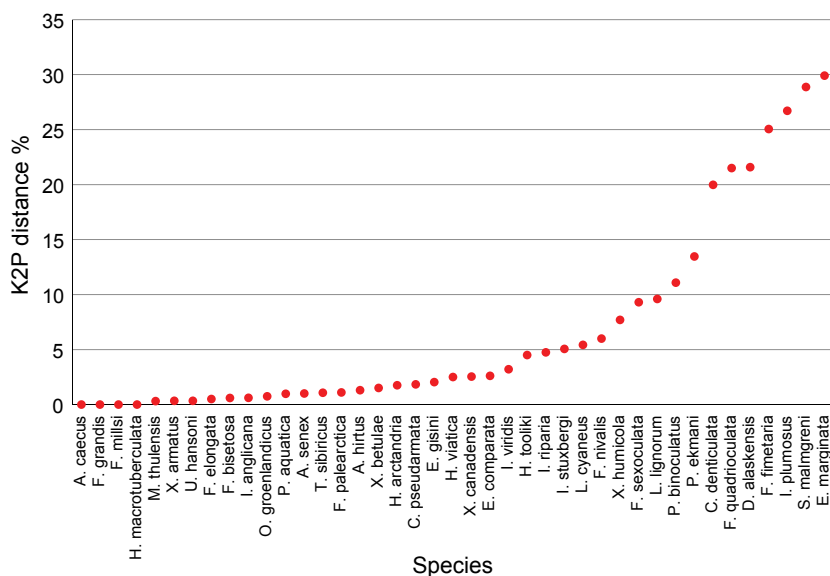


Fig. 2 Maximum intraspecific K2P divergences (%) for 41 identified species.

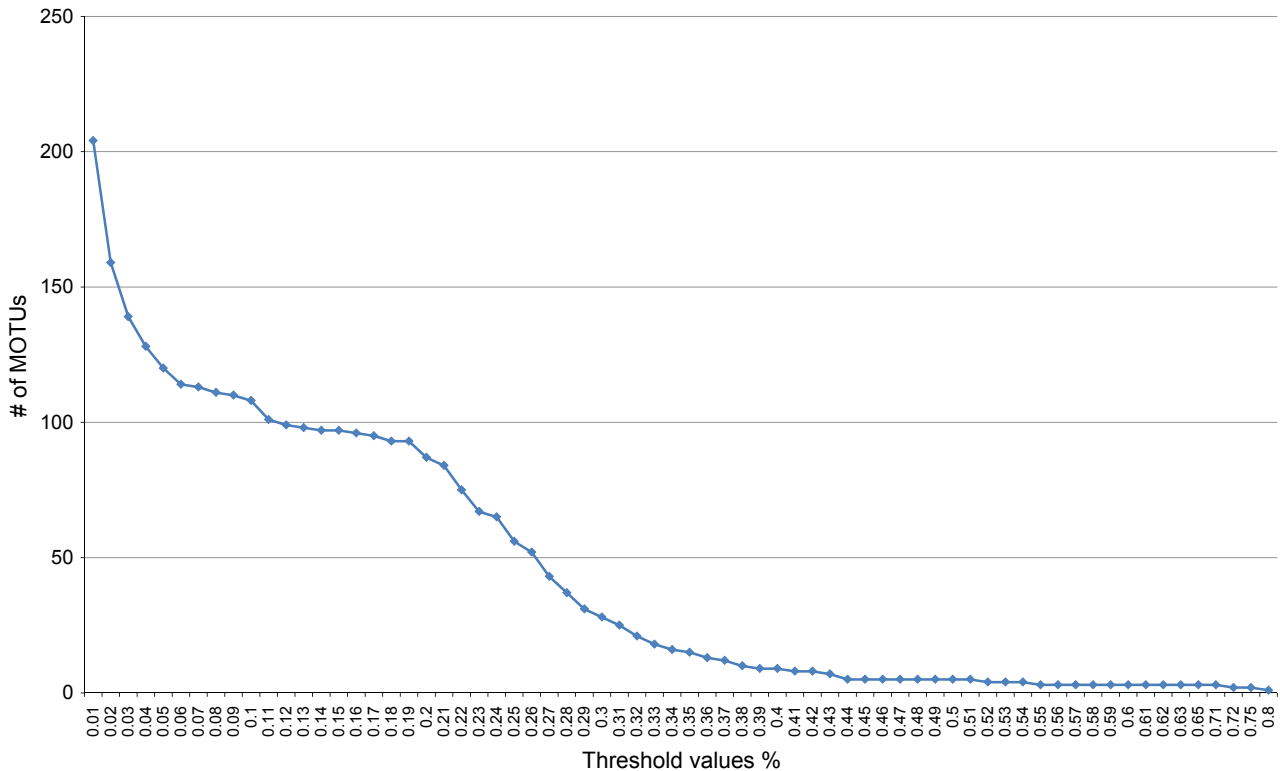


Fig. 3 Number of MOTUs plotted against different threshold values.

reached for the morphological and barcoding approaches (Fig. 5). The barcoding approach delivers both a higher count and estimation with Chao1 or ACE (Fig. 5). Concerning the four orders of Collembola, the inflection point was reached on the rarefaction curves calculated with molecular data, suggesting that sampling completion was nearly achieved and that the number of MOTUs observed is a close approximation to the total MOTU richness at Churchill. This was confirmed by both the Chao1 and ACE indices (Fig. 6), which showed that two more MOTUs can be expected in Poduromorpha, one more in Entomobryomorpha and four more in Symphypleona. No further progress is expected in the diversity of Neelipleona. The overall gain in diversity using barcodes was 8.4%, but when only morphologically identified species are considered, thus taking into account cases of cryptic diversity, the gain increased to 18%. This last estimate is more likely to reflect the actual benefit resulting from the use of DNA barcoding for the assessment of diversity.

The rarefaction curves for the different habitats (Fig. 7) showed that sampling completion was achieved for most. The absence/presence of the 97 MOTUs in the different habitats was surveyed (Table S3). The highest cumulative and estimated (Chao1 and ACE) richness values were observed in the tundra habitat, whereas the lowest values were found in the fen habitat (Fig. 7). The

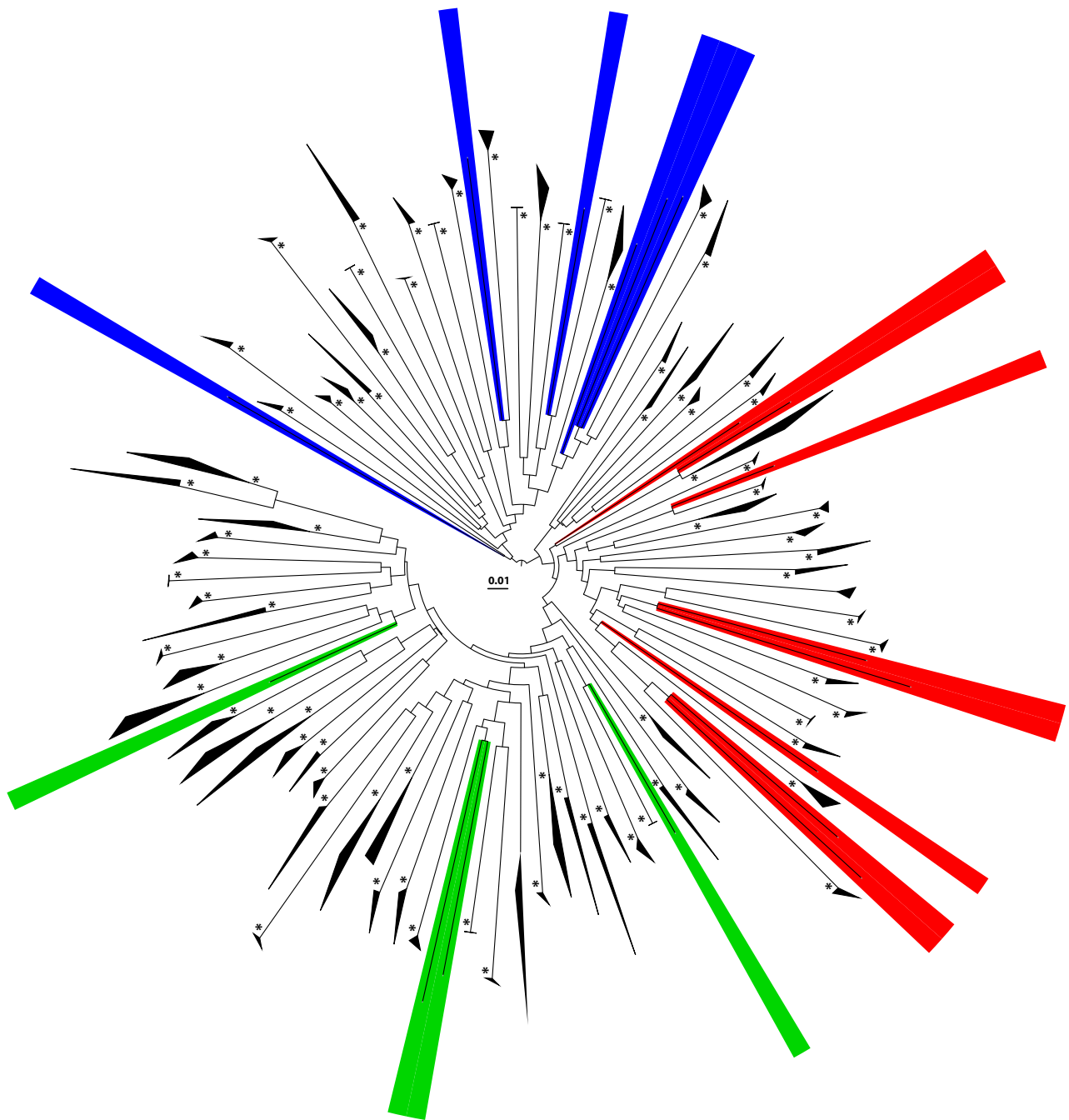
values for the Sørensen's index (Table 2) showed that the turnover in MOTUs between habitats was generally higher between coastal (beach, bluff, coastal tundra) and inland habitats (tundra, forest) than among habitats in each of these two categories. However, the values for the fen need to be interpreted with caution due to the small number of MOTUs in this habitat.

The first two axes of the correspondence analysis accounted, respectively, for 30.25% and 24.93% of the total inertia in the data set. Habitat ordination clearly separated coastal ecosystems (beach and coastal tundra, with negative scores on CA1 and CA2) from fen (positive scores on both CA1 and CA2) and forests (positive scores on CA1 and negative scores on CA2). Tundra and bluff habitats had somewhat intermediate positions (Figure 8a). Figure 8b and Table S4 indicate the species that characterized the assemblages in each of the different habitats.

## Discussion

### *Barcodes as an accurate species proxy – refining diversity estimation*

In the DNA barcode library assembled for this study, MOTUs were delineated with a 14% threshold criterion.



**Fig. 4** Neighbour-joining tree for the 97 MOTUs. Singletons present in the different orders are highlighted with colour on the branches: Blue = Symphypleona, 6 singletons, red = Poduromorpha, 8 singletons and green = Entomobryomorpha, 4 singletons. Nodes with bootstrap values  $\geq 99\%$  are marked with an asterisk.

The interspecific/inter-MOTUs divergences found in this study are consistent with those observed in prior studies on Collembola (Stevens *et al.* 2006; Rougerie *et al.* 2009; McGaughan *et al.* 2010; Porco *et al.* 2010b, 2013). One previous barcoding survey on Collembola from Arctic Canada (Hogg & Hebert 2004) reported that different species showed sequence divergences ranging from 8%

to 25%. The lower boundary of this range was established from two *Hypogastrura* species (8% between *H. sensilis* and *H. concolor*). This particular value is exceptionally low, as the mean interspecific divergence among the four species of *Hypogastrura* at Churchill was 24.1% (range = 20.9–26.8%), while mean intraspecific divergence was 0.62% (Range = 0–2.5%). This suggests

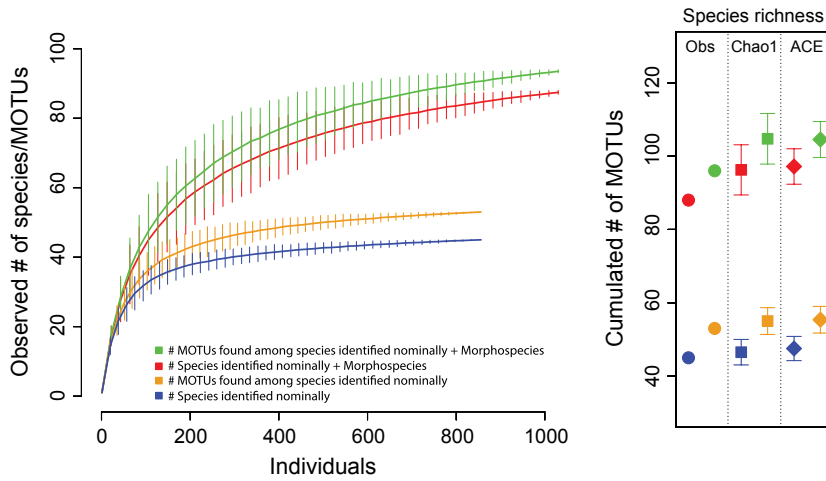


Fig. 5 Accumulation curves showing the increase in number of species/MOTUs with sampling effort both for the whole data set and for a data set restricted to the specimens identified to the species level. Observed (Obs) and estimated (Chao1 and ACE) species richness were plotted.

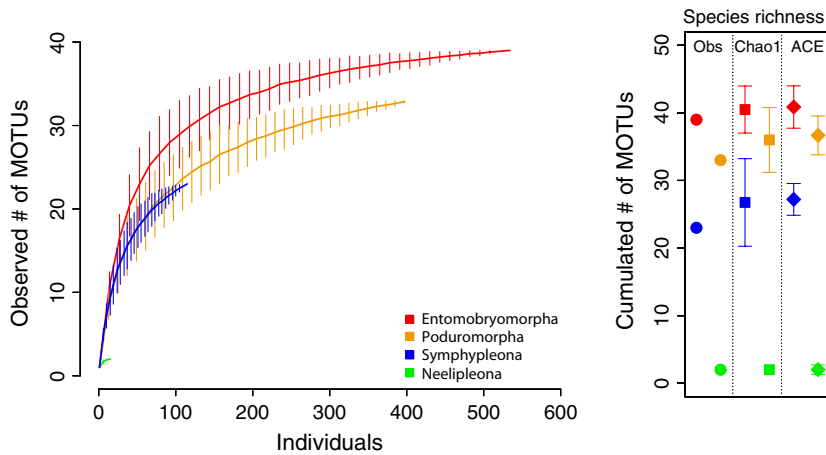


Fig. 6 Accumulation curves for 4 orders of Collembola showing the relationship between the number of MOTUs detected and sampling effort. Observed (Obs) and estimated (Chao1 and ACE) species richness were plotted.

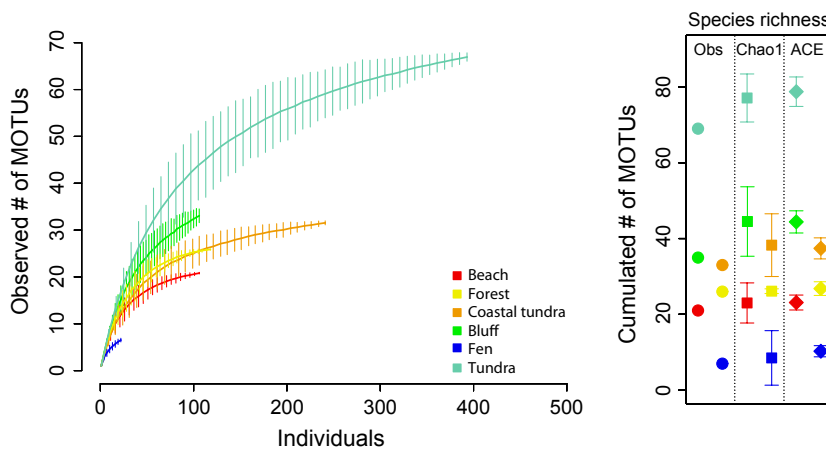


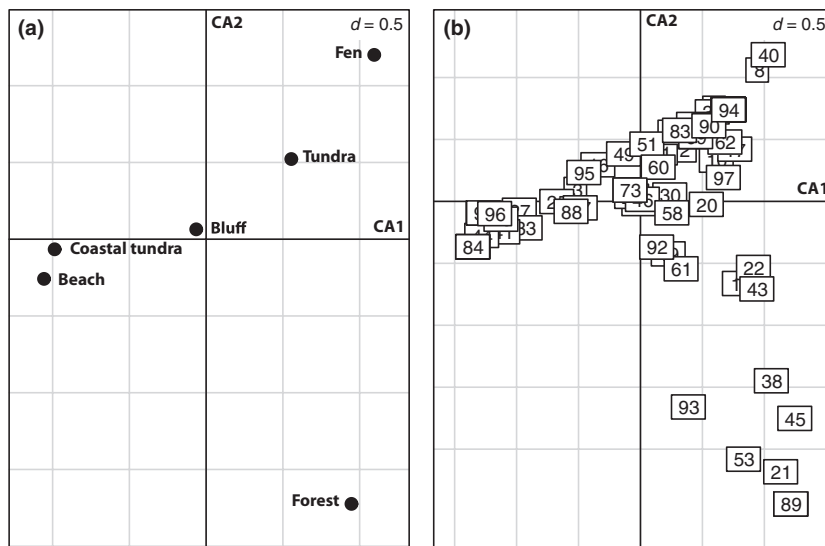
Fig. 7 Accumulation curves showing the increase in the number of MOTUs with sampling effort in the six habitat types. Observed (Obs) and estimated (Chao1 and ACE) species richness were plotted.

that these two species very close morphologically could be a single species, but this matter requires additional analyses. As a consequence, we adopted a conservative threshold of 14% as a basis for recognizing MOTUs in the Collembola at Churchill. This 14% threshold enabled

the exploration of the rest of the data set, which included taxa identified mostly to the genus level and in a few cases to the family level, permitting a comprehensive proxy estimate of collembolan diversity in the Churchill area and for the different types of habitats sampled.

**Table 2** Cumulated richness (CR), unique species (U. sp.) and beta diversity measures (Sorensen index, left diagonal/number of shared species, right diagonal)

	CR	U. sp.	Sorensen index/No. shared species					
			Beach	Bluff	Costal tundra	Tundra	Fen	Forest
Beach	21	5	–	16	12	11	1	7
Bluff	35	3	0.429	–	16	26	3	14
Costal tundra	33	7	0.556	0.529	–	20	2	10
Tundra	69	29	0.755	0.500	0.608	–	6	19
Fen	7	0	0.929	0.857	0.900	0.842	–	3
Forest	26	4	0.702	0.541	0.661	0.600	0.818	–

**Fig. 8** Results of the correspondence analysis of collembolan species assemblages: ordination of the habitats (a) and species/MOTUs (b) on the plan defined by the first two axes of the analysis (CA1 and CA2) accounting, respectively, for 30.25% and 24.93% of the total inertia of the data set. The numbers in (b) correspond to the number of the MOTUs as in Table S2.

Thirty-eight of the 45 species identified morphologically were assigned to a different MOTU, supporting the general equivalence between morphological species and MOTUs delimited with COI. However, the other seven species identified morphologically (*Ceratophysella denticulata*, *Desoria alaskensis*, *Entomobrya marginata*, *Folsomia fimetaria*, *Folsomia quadrioculata*, *Isotomurus plumosus* and *Sminthurides malmgreni*) showed high values of intraspecific divergence that reached interspecific levels. These high intraspecific divergences reflected the presence of discrete genetic entities (MOTUs) within the morphological boundaries of these nominal species. Similar cases have already been detected in other Collembola species and linked to cryptic diversity (Porco *et al.* 2012a,b). Accordingly, the nomenclature used for the designation of MOTUs detected in these previous surveys for *Podura aquatica* and *C. denticulata* complexes was applied here. Only one additional lineage of *C. denticulata* (*C. denticulata* L4) was detected at Churchill. For the 45 identified species, the detection of cryptic lineages led to a substantial rise in diversity for the Churchill area (18%). Moreover, the combination of our data set with the sequences

from the first barcoding survey in the Canadian Arctic (Hogg & Hebert 2004) revealed the presence of several additional cryptic lineages in four other species at Churchill (*P. aquatica*, *Sminthurides aquaticus*, *Entomobrya comparata* and *Sminthurides malmgreni*). As for the cases detected in our data set, mean interlineage distances (24.6%) and mean intralinea divergences (1.38%) corresponded to values typically found among well-defined species. Thus, these new lineages were named here following the same nomenclature used previously [(Porco *et al.* 2012a) and in this study]: *P. aquatica* L1 detected in Hogg & Hebert (2004) and Porco *et al.* (2012a) (L2 was found in this study), *S. aquaticus* L2 detected in Hogg & Hebert (2004) (L1 found in this study), *E. comparata* L2 detected in Hogg & Hebert (2004) (L1 found in this study), *S. malmgreni* L4 detected in Hogg & Hebert (2004) (L1, L2 and L3 found in this study). The lineage number was added to the species name accordingly in the corresponding GenBank, and BOLD records from Hogg & Hebert (2004). The consistent use of this notation will make it possible to track the distribution of the different MOTUs detected within a nominal species. This



brings to 10 the count of species exhibiting cryptic diversity in the Canadian Arctic habitat.

The use of DNA barcoding enables broad-ranging assessments of species richness when taxonomic expertise is unavailable or when a substantial fraction of the fauna is unknown, which occurs in many regions of the globe for Collembola (Deharveng 2004), as well as for various groups of soil invertebrates (Decaëns 2010). Moreover, the use of molecular taxonomy enables a finer resolution of species richness by revealing cryptic lineages in morphologically identified species. It is important to emphasize that all the cryptic lineages, except those in *C. denticulata*, exhibit sympatric distribution in some samples, supporting their likely specific status (Porco *et al.* 2012a).

#### *Diversity comparison with previous Arctic surveys*

The only previous survey for Collembola at Churchill (Hammer 1953) revealed the presence of 38 species versus the 97 MOTUs reported here. Interestingly, the latter number nearly matches the species count for Collembola for the whole Arctic of eastern North America (Babenko & Fjellberg 2006 – Table 3). Progress was made in this study in the evaluation of species richness in the different orders. The inflection point in the MOTUs rarefaction curves was reached, suggesting that little increase in species diversity will result from

further sampling of the four orders in the Churchill region (Fig. 6).

Hammer (Hammer 1953) did not detect any members of the order Neelipleona, while this study revealed the presence of one species. The number of MOTUs in Entomobryomorpha from Churchill is more than twice the number of species recorded by Hammer. In fact, it almost matches the number of species recorded in the whole Eastern Arctic of North America (Babenko & Fjellberg 2006). The diversity of MOTUs in Poduromorpha is a third higher than Hammer's survey (Hammer 1953), but the most spectacular increase involved the Symphypleona with 23 MOTUs in this study versus three species previously reported from Churchill (Hammer 1953) and nine species in the whole eastern Arctic of North America (Babenko & Fjellberg 2006) (Table 3). This number of MOTUs nearly matched the total diversity (36 species) of Symphypleona previously reported from the Arctic (Babenko & Fjellberg 2006). It opens exciting perspectives, which should be investigated further by a morphological examination of these MOTUs.

The much higher diversity of Symphypleona may be due, in part, to differences in collecting methods. The majority of the Symphypleona species were collected using sweep nets and pitfall traps, while Hammer (Hammer 1953) only used Berlese funnel extractions. However, the difficult taxonomy of this order and the very small number of specialists able to identify specimens in

**Table 3** Comparison of the diversity detected in this study with previous local and global surveys for families and orders of the class Collembola

Order	Family	This study	Hammer (1953)	Babenko & Fjellberg (2006) (Eastern America)	Babenko & Fjellberg (2006) (Global arctic)
Entomobryomorpha	Entomobryidae	7	1	5	23
	Isotomidae	31	15	38	132
	Tomoceridae	1	0	1	5
		39	16	44	160
Neelipleona	Neelidae	2	0	1	1
Poduromorpha	Hypogastruridae	11	7	24	67
	Neanuridae	11	6	16	53
	Odontellidae	1	1	1	4
	Onychiuridae	6	4	11	62
	Poduridae	1	0	1	1
	Tullbergiidae	3	2	3	23
		33	20	56	210
Symphypleona	Arrhopalitidae	4	0	1	5
	Bourletiellidae	5	0	1	11
	Katiannidae	2	0	1	8
	Sminthurididae	12	2	4	11
	Mackenziellidae	0	0	1	1
	Sminthurididae	0	0	0	8
	Dicyrtomidae	0	0	0	5
		23	2	8	36
Total		97	38	109	420

this group may have led to a higher level of cryptic diversity in the Symphypleona. This would account for the contrast between this study (24% of collembolan species in the Symphypleona), the Hammer survey (5% of collembolan species in the Smphypleona) and the whole Arctic region [9% of collembolan species belonging to Symphypleona – (Babenko & Fjellberg 2006)].

Six families were newly recorded for the Churchill region (Arrhopalitidae, Bourletiellidae, Katiannidae, Neelidae, Poduridae and Tomoceridae) (Table 3). In the other families, the number of MOTUs recorded in the present study was higher than the number of species in the previous survey by Hammer (1953) (Table 3). In some families, the number of MOTUs detected in the Churchill vicinity matched (Isotomidae) or even exceeded (Sminthurididae, Arrhopalitidae, Entomobryidae) the number of species previously recorded from the whole of eastern Arctic of North America (Table 3). In Sminthurididae, the number of MOTUs even exceeded the total species found in the whole of the Arctic (Babenko & Fjellberg 2006). Two MOTUs of Neelidae were found, while only one species was recorded in the previous Arctic surveys for this order.

Based on the presumption that MOTUs are species, if the increase in diversity reported in this study compared with the earlier investigation (Hammer 1953) is extrapolated to the eastern Arctic of North America area as defined by Babenko & Fjellberg (2006), the diversity in this region would reach 278 species instead of 109. An extrapolation to the whole Arctic biome would push the count from 420 to 1071 species. These estimates are likely high as Churchill is at the southern limit of the Canadian Arctic, and species richness might be higher than in the high Arctic. However, the rough estimates, based on our survey of this small region of Arctic Canada, suggest much higher species diversity than expected from prior work. This result will change the perception of the diversity of the group in this habitat considered poor and lead to question the gap between observed and actual diversity in many other environments.

#### *Diversity among the different habitats in Churchill*

Six habitats were sampled at Churchill (coastal tundra, tundra, beach, bluff, forest and fen). The rarefaction curves and MOTUs richness estimators calculated for these different habitats showed that the inflection point was reached for most habitats: little further increase in diversity is likely to be revealed by further sampling (maximum estimated increase: coastal tundra 15%, tundra 13%, forest 0%, beach 9% – Fig. 7), while fen and bluff should yield a higher diversity than observed in our study (maximum estimated increase, respectively, of 42% and 31% – Fig. 7).

Forty-eight of the MOTUs (49.5%) found in this study were specific to one of the six habitat types, but most (40) were only known from one population, sometimes only from one individual (14) (Table S3, Fig. 4). The majority of these habitat-specific MOTUs were found in the most species-rich habitat, the tundra, where 42% of the collected MOTUs are unique to this habitat. This high specificity of the species assemblage in the tundra is likely due to the fact that this habitat is constituted of a mosaic of different vegetation communities (Collinson 1988), thus presenting higher numbers of niches and trophic resources. The other extremes are the fen with no unique MOTUs, and the bluff with only 8.5% unique MOTUs which are much more homogeneous habitats with fewer niches and lower resources. Beach and coastal tundra have over 20% unique MOTUs, while the forest has 15%. Only 25.7% of the MOTUs were present in more than two habitats and only *Folsomia nivalis* cited as one of the most common species in North America and also detected in Far East Russia was found in all habitats. Three species/MOTUs were observed only in coastal habitats: *H. arctandria*, *H. viatica* and *I. plumosus* L1. The correspondence analysis highlights clear ecological structuring of collembolan species assemblages in the Churchill area. For instance, coastal habitats (i.e. beach, bluff and coastal tundra) are all clustered together on the factorial plan defined by the first two axes of the CA (Fig. 8). This conclusion is further supported by the beta diversity measurements among these habitats which indicate low turnover between their communities (Table 2), with the sharing of many species. This could be explained by the geographical proximity of these coastal habitats, but also by their similarities in climatic conditions and trophic resources. They are clearly separated from inland habitats on CA1, suggesting that the distance from the coast could represent the first structuring factor of collembolan species assemblages at the study site. The fen and the forest habitats are further opposed on CA2 and also exhibit the highest values of Sørensen's index, showing a strong influence of inland vegetation cover on assemblage composition and high specificity in the composition of these communities.

Except for the lineages of *S. malmgreni*, which were all found in the tundra, the different cryptic lineages revealed in morphologically identified species occurred in different types of habitats (Table S3). This suggests dissimilarities in both the biology and ecological requirements of these lineages as already suspected in other cryptic species complexes of *Collembola* (Porco *et al.* 2012a,b). This observation also supports the species level of these lineages, emphasizing the importance of taking this diversity into account.

### Introduced species and conservation

As the seaport at Churchill gained functionality in 1931, millions of tons of cargo, mostly wheat, have transited through it, making this a major seaport for northern Canada. This high volume of ship traffic to other regions of North America, as well as Asia, Europe and Africa (Chan *et al.* 2011), has likely led to the repeated introduction of exotic species. The new 'Arctic Gateway' project will link Churchill to Murmansk (<http://www.arcticbridge.com/>) aided by the warming of Arctic waters generating a longer ice-free season for navigation (Comiso *et al.* 2008; Galbraith & Larouche 2011). If no preventive steps are taken, this endeavour could threaten local diversity through the introduction of faunal elements from the Russian Arctic leading to the homogenization of the species assemblages across the Arctic. Furthermore, rising temperatures will directly impact species adapted to colder conditions reducing their distribution (Campbell *et al.* 2009), while favouring the introduction of potentially competitive southern species. All these factors suggest the possibility of an invasion meltdown in Arctic ecosystems. A recent survey of human-mediated introductions of soil invertebrates in another sub-Arctic location, the Svalbard Island, supports this scenario (Coulson *et al.* 2013).

Our data set shows examples of fauna exchange for Collembola. For instance, one of the MOTUs detected in the present survey, *C. denticulata* L3, is likely introduced as it has also been detected in South Africa, Australia, New Zealand (Porco *et al.* 2012a) and France (unpublished data). Its broad distribution can only be explained by passive dispersal through human activities. Moreover, *C. denticulata* L3 was only detected near the Churchill grain elevator (highly disturbed habitat), while *C. denticulata* L4 was present from the rest of the coast to deep inland up to Twin Lakes (bluff, forest, tundra). It is noteworthy that *C. denticulata* L4 has not yet been detected in any of the countries where *C. denticulata* L3 was found. Considering both MOTUs as a single species would have led to erroneous considerations concerning their distribution and invasive status (i.e. presence in disturbed and natural habitats). Introduction can also be suspected for *P. aquatica* L2 as it was only recovered from a disturbed site (in the vicinity of the grain elevator) although numerous favourable other ponds were sampled without success in other areas near Churchill. By contrast, the European species *E. multifasciata* was only found once in the coastal tundra habitat, far from human disturbance, suggesting its acclimation to undisturbed environments. The occurrence of such cases emphasizes further the need for extensive baseline surveys to monitor new species introductions and invasions in the communities.

### Conclusion

Our study exemplifies the usefulness of DNA barcoding for estimating and reassessing species diversity in groups of soil invertebrates heavily impacted by the taxonomic impediment. An assessment of collembolan diversity was gained through the use of a 14% sequence threshold to delineate MOTUs. This species proxy approach made it possible to examine essential research themes, such as the comparison of communities among habitats and the monitoring of introduced species. The present survey may enable further studies investigating the dynamic of collembolan communities in the Churchill area, as well as the impacts of species introductions and climate warming. Besides, the use of this approach allowed a more comprehensive assessment of the diversity through the detection of unexpectedly high cryptic diversity in a habitat considered as species poor for Collembola. These results emphasize the value of a combined morphological-DNA approach to document and compare accurately diversity among biomes, a much needed task for monitoring of the impacts of global change on ecosystems.

### Acknowledgements

This study was supported by grants to PDNH from NSERC and from the government of Canada through Genome Canada and the Ontario Genomics Institute in support of the International Barcode of Life Project. We thank the Churchill Northern Studies Centre and the expedition team from the Biodiversity Institute of Ontario for aid with logistics and sampling. We also thank Serguei Ponomarenko for samples from more distant sites in the Churchill region, Aurélie Husté for the help with cartography, Feng Zhang for scanning many precious bibliographic references and Ian Hogg who authorized our nomenclature of cryptic lineages to be extended to his Arctic project.

### References

- Babenko AB, Fjellberg A (2006) *Collembola Septentrionale: A Catalogue of Springtails of the Arctic Regions*. KMK Scientific Press, Russia.
- Benzécri JP (1973) *L'analyse des données. II L'analyse des correspondances*. Bordas, Paris.
- Bretfeld G (1999) *Symphyleona*, Staatliches Museum für Naturkunde Görlitz edn. Staatliches Museum für Naturkunde Görlitz.
- Campbell A, Kapos V, Scharlemann JPW *et al.* (2009) Review of the literature on the links between biodiversity and climate changes: impacts, adaptation and mitigation. *CBD Technical Series*, 42, 1–124.
- Chan FT, Bronnenhuber JE, Bradie JN *et al.* (2011) *Risk Assessment for Ship-Mediated Introductions of Aquatic Nonindigenous Species to the Canadian Arctic* Fisheries and Oceans Canada, Research Document 2011/105, (ed. Secretariat FaOC-DCSA), pp. 1–93.
- Chessel D, Dufour AB (2004) The ade4 package -I- One-table methods. *R News*, 4, 5–10.
- Christiansen K, Bellinger P (1998). *The Collembola of North America north of the Rio Grande. A taxonomic analysis*. Grinnell College, Grinnell, IA.
- Cicconardi F, Nardi F, Emerson BC, Frati F, Fanciulli PP (2010) Deep phylogeographic divisions and long-term persistence of forest

- invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. *Molecular Ecology*, **19**, 386–400.
- Collinson AS (1988) *Introduction to World Vegetation*, 2nd edn. Academic Division of Unwin Hyman Ltd, London.
- Comiso JC, Parkinson CL, Gersten R, Stock L (2008) Accelerated decline in the Arctic Sea ice cover. *Geophysical Research Letters*, **35**, 1–6.
- Coulson SJ, Fjellberg A, Gwiazdowicz DJ *et al.* (2013) Introduction of invertebrates into the High Arctic via imported soils: the case of Barentsburg in the Svalbard. *Biological Invasions*, **15**, 1–5.
- Decaens T (2010) Macroecological patterns in soil communities. *Global Ecology and Biogeography*, **19**, 287–302.
- Deharveng L (2004) Recent advances in Collembola systematics. *Pedobiologia*, **48**, 415–433.
- Fjellberg A (1980) *Identification Keys to Norwegian Collembola*. Norsk Entomologisk Forening, As, Norway.
- 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.
- Galbraith PS, Larouche P (2011) Sea-surface temperature in Hudson Bay and Hudson Strait in relation to air temperature and ice cover breakup, 1985–2009. *Journal of Marine Systems*, **87**, 66–78.
- Gotelli NJ, Entsminger GL (2006) *EcoSim: Null Models Software for Ecology*. Acquired Intelligence and Kesey-Bear, Jericho, VT, USA.
- Greenacre MJ (1984) *Theory and Applications of Correspondence Analysis*. Academic Press, London, Orlando, etc.
- Hajibabaei M, DeWaard JR, Ivanova NV *et al.* (2005) Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **360**, 1959–1967.
- Hajibabaei M, Smith MA, Janzen DH *et al.* (2006) A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, **6**, 959–964.
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.
- Hammer M (1953) Collembolans and Oribatids from the Thule district (North West Greenland) and Ellesmere Island (Canada) collected by J. C. Troelsen and Chr. Vibe. *Meddelelser om Gronland*, **136**, 1–16.
- Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, 313–321.
- Hogg ID, Hebert PDN (2004) Biological identification of springtails (Hexapoda: Collembola) from the Canadian Arctic, using mitochondrial DNA barcodes. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **82**, 749–754.
- Hopkin SP (1997) *Biology of the Springtails (Insecta: Collembola)*. Oxford University Press, New York, Tokyo.
- Ivanova NV, Dewaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, **6**, 998–1002.
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.
- Lemon J (2006) Plotrix: a package in the red light district of R. *R-News*, **6**, 8–12.
- Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, **23**, 127–128.
- Mari Mutt JA, Bellinger PF (1990) A catalog of the Neotropical Collembola, including Nearctic areas of Mexico. *Flora & Fauna Handbook*, i–xii, 1–237.
- Mari Mutt JA, Bellinger PF (1996) Supplement to the Catalog of the Neotropical Collembola – August 1989 to April 1996. *Caribbean Journal of Science*, **32**, 166–175.
- May RM (2010) Tropical Arthropod Species, More or Less? *Science*, **329**, 41–42.
- McGaughan A, Torricelli G, Carapelli A *et al.* (2010) Contrasting phylogeographical patterns for springtails reflect different evolutionary histories between the Antarctic Peninsula and continental Antarctica. *Journal of Biogeography*, **37**, 103–119.
- Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How Many Species Are There on Earth and in the Ocean? *PLoS Biology*, **9**, e1001127, 1001121–1001128.
- Oksanen J, Blanchet G, Kindt R *et al.* (2012) *Vegan: Community Ecology Package*. R package version 2.0-3.
- Plaisance L, Caley MJ, Brainard RE, Knowlton N (2011) The Diversity of Coral Reefs: what Are We Missing? *PLoS ONE*, **6**, e25026.
- Porco D, Rougerie R, Deharveng L, Hebert P (2010a) Coupling non-destructive DNA extraction and voucher retrieval for small soft-bodied Arthropods in a high-throughput context: the example of Collembola. *Molecular Ecology Resources*, **10**, 942–945.
- Porco D, Bedos A, Deharveng L (2010b) Description and DNA barcoding assessment of the new species *Deutonura gibbosa* (Collembola: Neanuridae: Neanurinae), a common springtail of Alps and Jura. *Zootaxa*, **2639**, 59–68.
- Porco D, Bedos A, Greenslade P *et al.* (2012a) Challenging species delimitation in Collembola: cryptic diversity among common springtails unveiled by DNA barcoding. *Invertebrate Systematics*, **26**, 470–477.
- Porco D, Potapov M, Bedos A *et al.* (2012b) Cryptic Diversity in the Ubiquitous Species *Parisotoma notabilis* (Collembola, Isotomidae): a Long-Used Chimeric Species? *PLoS ONE*, **7**, e46056.
- Porco D, Decaens T, Deharveng L *et al.* (2013) Biological invasions in soil: DNA barcoding as a monitoring tool in a multiple taxa survey targeting European earthworms and springtails in North America. *Biological Invasions*, **15**, 899–910.
- Potapov M (2001) Synopses on Palaearctic Collembola. Volume 3. Isotomidae. *Abhandlungen und Berichte des Naturkundemuseums Goerlitz*, **73**, 1–603.
- R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rougerie R, Decaens T, Deharveng L *et al.* (2009) DNA barcodes for soil animal taxonomy. *Pesquisa Agropecuaria Brasileira*, **44**, 789–802.
- Saitou N, Nei M (1987) The neighbor-joining method – a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.
- Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, **75**, 7537–7541.
- Stevens MI, Greenslade P, Hogg ID, Sunnucks P (2006) Southern hemisphere springtails: could any have survived glaciation of Antarctica? *Molecular Biology and Evolution*, **23**, 874–882.
- Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

---

D.P. carried out the field sampling and managed the molecular analyses. D.P., D.S. and L.D identified the specimens. D.P. and T.D. analysed the data. D.P. wrote the manuscript and D.S., L.D., T.D. and P.H. participated in finalizing the text. P.H. obtained funding and managed the overall Churchill programme.

---

### Data Accessibility

DNA sequences: BOLD project <http://dx.doi.org/10.5883/DATASET-CHU-COL1>.

GenBank Accessions: GU657078–GU657335, HM390591–HM390627, HM390653–HM390662, HM424138–HM424148, HM431644–HM431645, HM893770–HM893811, HQ992117, JF884220–JF884222, JF884258, JN269617–JN269618, JN306349–JN306356, JX261789, JX261792, JX261796, JX261798, JX261801, JX261804, JX261808, JX261809, JX261814, JX261819, JX261833, JX261837, JX261843, JX261844, JX261848, JX261857, JX261876, JX261884, JX261886, KF641929–KF642599.

### Supporting Information

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

**Fig. S1** Variation of the intraspecific/intraMOTUs and interspecific/interMOTUs divergence for K2P distances (%).

**Fig. S2** Neighbour-Joining tree for the 1066 specimens sequenced for COI. Orders are highlighted with colour on the labels of the branches: Blue = Symphypleona, Red = Poduromorpha and Green = Entomobryomorpha. Intraspecific variation is displayed as an histogram above the branch labels for each MOTU.

**Table S1** MOTUs/species checklist, sampling localities and collecting effort (number of specimens and populations sampled).

**Table S2** Mean intraspecific/intraMOTUs and interspecific/interMOTUs divergence for K2P distances (%).

**Table S3** Presence/absence of the MOTUs, count of the number of specimens and populations sampled in the six habitats.

**Table S4** Species scores on the first two axes of the correspondence analysis.