Biological identification of springtails (Hexapoda: Collembola) from the Canadian Arctic, using mitochondrial DNA barcodes

Ian D. Hogg and Paul D.N. Hebert

Abstract: We evaluated sequence diversity in the mitochondrial cytochrome-*c* oxidase I (COI; EC 1.9.3.1) gene as a tool for resolving differences among species of Arctic springtails. The Collembola examined in this analysis were collected from Igloolik, Cornwallis, and Somerset islands and included representatives from all major families found in the Arctic. Members of 13 genera and 19 species were examined, including 4 species of the genus *Folsomia* and 3 species of the genus *Hypogastrura*. In all cases, species were successfully discriminated. Sequence divergences within species were generally less than 1%, whereas divergences between species were greater than 8% in all cases. Divergences among individuals of one species of *Folsomia* were much higher (up to 13%), but this likely represents the presence of an undescribed sibling species. We conclude that DNA barcoding is a powerful tool for identifying species of Collembola and should regularly be useful as a complement to traditional, morphological taxonomy.

Résumé : Nous avons évalué l'utilisation de la diversité des séquences du gène mitochondrial de la cytochrome-*c* oxydase I (COI; EC 1.9.3.1) comme outil pour déterminer les différences spécifiques chez des collemboles arctiques. Les collemboles étudiés dans cette analyse proviennent des îles Igloolik, Cornwallis et Somerset et comprennent des représentants de toutes les familles importantes présentes dans l'arctique. Nous avons examiné des spécimens de 13 genres et 19 espèces, y compris quatre espèces du genre *Folsomia* et trois espèces du genre *Hypogastrura*. Dans tous les cas, il a été possible de reconnaître les espèces avec succès. Les divergences entre les séquences dans une même espèce sont généralement de moins de 1 %, alors que les divergences entre les espèces sont supérieures à 8 % dans la plupart des cas. Les divergences parmi les individus d'une des espèces de *Folsomia* est beaucoup plus élevée (jusqu'à 13 %), mais il y a probablement présence d'une espèce soeur non décrite. En conclusion, le codage ADN est un outil puissant pour l'identification des espèces de collemboles et il devrait normalement s'avérer utile comme méthode complémentaire à la taxinomie morphologique traditionnelle.

[Traduit par la Rédaction]

Introduction

The ability to accurately identify species is fundamental to ecological research, particularly studies of comparative ecology and biological diversity. Unfortunately, for a growing number of groups, the chronic underfunding of traditional taxonomy (Daugherty et al. 1990; Savage 1995) has led to a serious shortfall in the experts needed to carry out identifications. The use of DNA-based technologies has been suggested as the best option to "bridge" the gap between available taxonomic expertise and the need for an identification capability (Tautz et al. 2003). Although this approach is not without controversy (e.g., Lipscomb et al. 2003), the taxonomic impediment is now serious enough that it is critical to seek a novel solution. Hebert et al. (2003*a*) have proposed that the analysis of sequence diversity in short stretches of mitochondrial DNA — in particular, the cytochrome-*c* oxidase I (COI) gene — can provide an effective tool for species diagnosis in the animal kingdom. In fact, they argued that sequence diversity in this gene could be used to create a "barcoding" system that would enable the identification of all animal life. The potential success of this endeavour was supported by an analysis showing that deep genetic divergences were the rule for 13 000 closely allied species from a range of animal phyla (Hebert et al. 2003*b*). However, no taxonomic group except the Lepidoptera (Hebert et al. 2003*a*) has been examined in detail.

In this paper, we examine the utility of COI barcoding for the discrimination of collembolan species. We targeted this group for analysis because its members are among the most diverse and numerically dominant of soil arthropods (Petersen and Luxton 1982). Our work focuses on the analysis of Collembola from the Canadian Arctic, a region that supports a diverse fauna of these organisms (Danks 1981). Despite this diversity, the distribution and taxonomy of Arctic Collembola is poorly known (Fjellberg 1986; Babenko 1994). By contrast, their Antarctic counterparts have received considerably more attention (see Hogg and Stevens 2002 for review). This imbalance of effort is unfortunate given the much greater species diversity of Collembola in

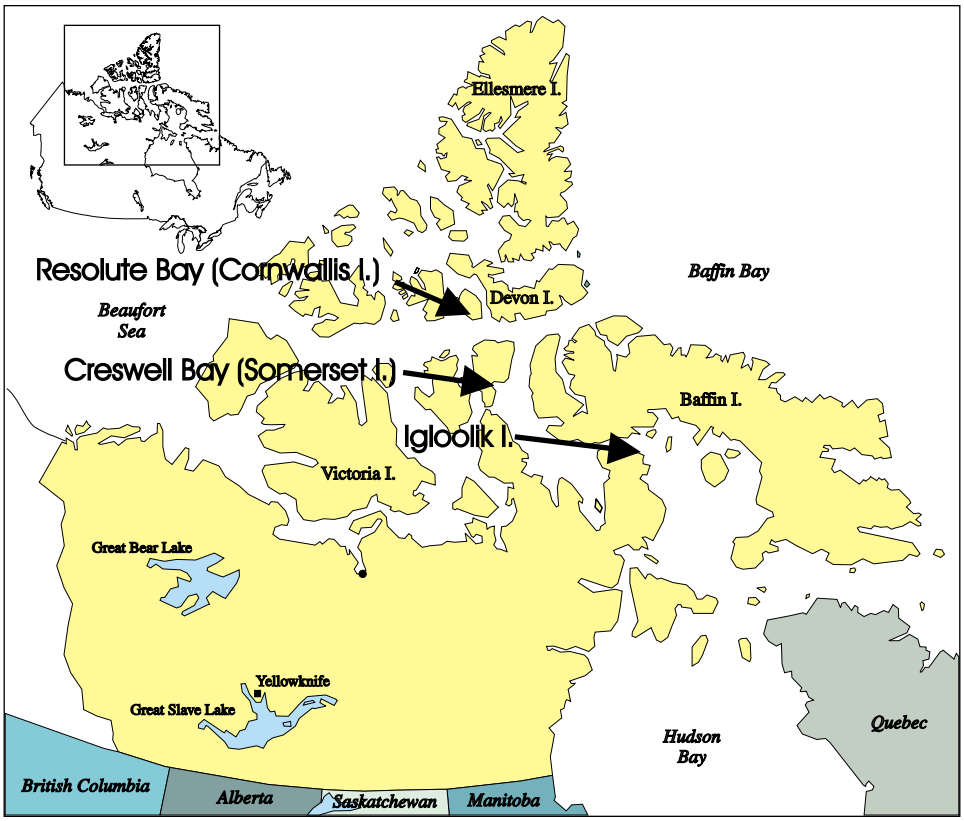
Received 19 August 2003. Accepted 24 March 2004. Published on the NRC Research Press Web site at http://cjz.nrc.ca on 13 July 2004.

I.D. Hogg.[1] Centre for Biodiversity and Ecology Research, Department of Biological Sciences, University of Waikato, Private Bag 3105, Hamilton, New Zealand.
P.D.N. Hebert. Department of Zoology, University of Guelph, ON N1G 2W1, Canada.

[1]Corresponding author (e-mail: hogg@waikato.ac.nz).

 doi: 10.1139/Z04-041

Fig. 1. Map of study area showing sampling locations on Cornwallis, Somerset, and Igloolik islands.



the Arctic. However, one major barrier impedes future work on the Arctic fauna: species identifications are often challenging and require considerable taxonomic expertise. Accordingly, Arctic Collembola are a group for which a DNA barcoding system would find immediate application.

Previous information on mtDNA diversity in Collembola has been restricted to the Antarctic (e.g., Frati et al. 2000*b*; Stevens and Hogg 2003) and European (Carapelli et al. 1995*a*, 1995*b*) faunas. Only limited data exist for the COI gene (Frati et al. 2000*a*; Stevens and Hogg 2003). The present analysis extends the available data substantially and examines whether COI sequence diversity successfully resolves species-level differences among the dominant Collembola from Arctic Canada.

Methods

Study sites and collection of animals

Samples were collected in August 2001 from two main areas, Igloolik (Igloolik Island) and Resolute Bay (Cornwallis Island), Nunavut, Canada (Fig. 1). However, additional collections were made on Somerset Island (Creswell Bay) and on Devon Island, for a total of 56 sampling sites. Samples were obtained from a variety of habitats including soil, moss, vascular plants, stream drift, lake margins, and lake strandline detritus. Animals were extracted using a modified Berlese funnel and stored in 100% ethanol until used for genetic and morphological analyses. Where possible, efforts were made to include taxa that were found at more than one location (based on personal observation and Fjellberg 1986), as well as those that were numerically dominant. Individuals were initially identified using available taxonomic keys (e.g., Folsom 1937; Christiansen and Bellinger 1980) and primary references (e.g., Babenko 1994). Their identity was subsequently confirmed by an internationally recognised expert (A. Fjellberg) in collembolan taxonomy. A total of 19 species were recognized from 11 sites and selected for analysis (Table 1), providing taxonomic coverage of all major Arctic families. Fjellberg (1986), in his review of Arctic Collembola, reported approximately 50 species from the Canadian high Arctic. Accordingly, the 19 species identified and analysed in our study represent a significant proportion of the fauna and include most of the common taxa in this region.

Genomic DNA was extracted from whole-body homogenates using proteinase K methods. A 710-bp region of the mitochondrial COI gene was amplified by polymerase chain reaction (PCR) from the DNA of at least three individuals from each putative species using the primers LCO1490 and HCO2198 (Folmer et al. 1994). The 50-µl polymerase chain reactions contained 1 µL of DNA template, 5.0 µL of 10× PCR buffer (Boehringer-Mannheim), 0.2 µM of each primer, 2.2 mM $MgCl_2$, 0.2 mM of dNTP, and 1 unit (1 U ≈ 16.67 nkat) of *Taq* DNA polymerase (Roche). The PCR conditions consisted of 1 min at 94 °C; five cycles of 1 min at 94 °C, 1 min at 45 °C, and 1 min at 72 °C; 35 cycles of 1 min at 94 °C, 1 min at 51 °C, and 1 min at 72 °C; and 5 min at 72 °C. PCR products were cleaned using gel purification (QIAEX II gel extraction kit, QIAGEN Inc.) and sequenced in one direction using primer LCO1490 on an ABI

Table 1. Study taxa, number of individuals included in the analysis (*n*), sample location, habitat type, and distribution in the Canadian Arctic.

Species	*n*	Sample location			Habitat type	Distribution*
		Site	Island	Coordinates		
Archisotoma polaris Fjellberg and Poinsot, 1975	3	R9	Somerset	74°10N, 94°12W	Intertidal zone	E,S
Desoria tschernovi Martynova, 1974	1	R11	Somerset	72°44.260N, 94°26.147W	Wet moss	S
Entomobrya comparata Folsom, 1919	3	I16	Igloolik	69°20.598N, 81°39.454W	Rock	D,E,I
Folsomia bisetosa Gisin, 1953	3	R2	Cornwallis	74°40.465N, 94°55.328W	Moss	C,D,E
Folsomia quadrioculata (Tullberg, 1871)	3	R11	Somerset	72°44.260N, 94°26.147W	Wet moss	B,C,D,E,S
Folsomia regularis Hammer, 1953	3	R6	Cornwallis	75°23.147N, 94°43.894W	Moss	B,C,D,E,K,R
Folsomia sexoculata (Tullberg, 1871)	3	R2	Cornwallis	74°40.465N, 94°55.328W	Moss	C,D
Hypogastrura concolor (Carpenter, 1900)	3	R21	Cornwallis	74°42.634N, 95°02.276W	Plant (angiosperm)	C,D,E
Hypogastrura sensilis (Folsom, 1919)	3	I20	Igloolik	69°20.455N, 81°48.611W	Rock	C,D,E,I,K
	2	R2	Cornwallis	74°40.465N, 94°55.328W	Moss	
Hypogastrura sp.	2	I14	Igloolik	69°21.117N, 81°51.810W	Lake strandline	C,I
	1	R2	Cornwallis	74°40.465N, 94°55.328W	Moss	
Isotoma anglicana Lubbock, 1862	2	R2	Cornwallis	74°40.465N, 94°55.328W	Rock	C
Isotomurus sp. (near *palustroides*/*stuxbergi*)	2	I14	Igloolik	69°21.117N, 81°51.810W	Lake strandline	I
Morulina mackenziana Hammer, 1953	1	I10	Igloolik	69°22.051N, 81°48.065W	Rock	D,E,I
Onychiurus groenlandicus (Tullberg, 1876)	3	R2	Cornwallis	74°40.465N, 94°55.328W	Moss	B,C,D,E,K,R
Podura aquatica L., 1758	3	I14	Igloolik	69°21.117N, 81°51.810W	Lake strandline	B,C,D,E,I
	1	R6	Cornwallis	75°23.147N, 94°43.894W	Moss	
Stachanorema tolerans Babenko, 1994	3	R11	Somerset	72°44.260N, 94°26.147W	Wet moss	D,S
Sminthurides aquaticus (Bourlet, 1843)	1	I9	Igloolik	69°22.151N, 81°48.141W	Stream drift	E,I
Sminthurides malmgreni (Tullberg, 1876)	2	I10.2	Igloolik	69°22.051N, 81°48.065W	Stream drift	C,D,E,I
	1	R7	Cornwallis	75°22.682N, 94°00.452W	Lake strandline	
Vertagopus brevicaudus (Carpenter, 1900)	2	R7	Cornwallis	75°22.682N, 94°00.452W	Wet moss	C,K
	3	R21	Cornwallis	74°42.634N, 95°02.276W	Plant (angiosperm)	

*Canadian Arctic distribution records are based on Fjellberg (1986), Babenko (1993), and I.D. Hogg and P.D.N. Hebert (unpublished data), and they include Bathurst (B), Cornwallis (C), Devon (D), Ellesmere (E), Igloolik (I), King Christian (K), Ellef Ringnes (R), and Somerset (S) islands.

377 automated DNA sequencer at the University of Guelph. The accuracy of sequencing was verified by duplicate sequencing of DNA from selected individuals. All sequences were verified as being derived from Arthropoda using the GenBank BLAST algorithm. The resulting sequences were edited and aligned using the program Sequencher (version 4.1; Gene Codes Corp.). Nucleotide divergences within and between species were calculated using uncorrected *p* distances. A multidimensional scaling analysis was used to provide a graphical summary of species similarities in two-dimensional, Euclidean space (Lessa 1990). A neighbour joining analysis (performed using PAUP* version 4.0*b*10, Swofford 2002) was used to examine relationships among the taxa. All sequences obtained in this study have been submitted to GenBank (accession Nos. AY665303–AY665356).

Results

A 634-bp fragment of the COI gene was sequenced from a total of 54 individuals. Alignment begins at position 1548 when aligned with the published *Drosophila yakuba* sequence (Folmer et al. 1994), which was retrieved from GenBank (accession No. X03240). No insertions or deletions were detected in any of the sequences. Nucleotide composition averaged over all taxa showed an A–T bias (A = 26%, T = 34%, C = 21%, G = 19%). Base frequencies were homogeneous among sequences (χ^2_{156} = 131.41, P = 0.92).

Each of the species that we examined possessed a unique array of COI sequences with low divergences among conspecific individuals (<1% in most cases) (Table 2). Exceptions were single specimens of *Sminthurides malmgreni* and *Folsomia quadrioculata* that showed 5% and 13% divergence, respectively, from other individuals at the same location. Even in these two exceptional cases, the divergent individuals still grouped with other members of their nominate species (Fig. 2). By contrast, divergences among species exceeded 8% in all cases and were >15% among 18 of the 19 species. The most divergent taxon was *Podura aquatica*, which had a mean sequence divergence of >25% (range 23%–28%) from the other taxa (Table 2). Multidimensional scaling analysis showed that individuals of the same species always grouped closely together. In fact, individuals of *Archisotoma polaris*, *Entomobrya comparata*, *Hypogastrura concolor*, and *P. aquatica* had identical sequences (Fig. 2). Four species were collected on both Cornwallis and Igloolik islands: *Hypogastrura tullbergi*, *Hypogastrura sensilis*, *P. aquatica*, and *S. malmgreni*. These islands are separated by >600 km, but all four species showed low sequence divergence between the two regions (<1% in all cases) and the *P. aquatica* sequences were identical.

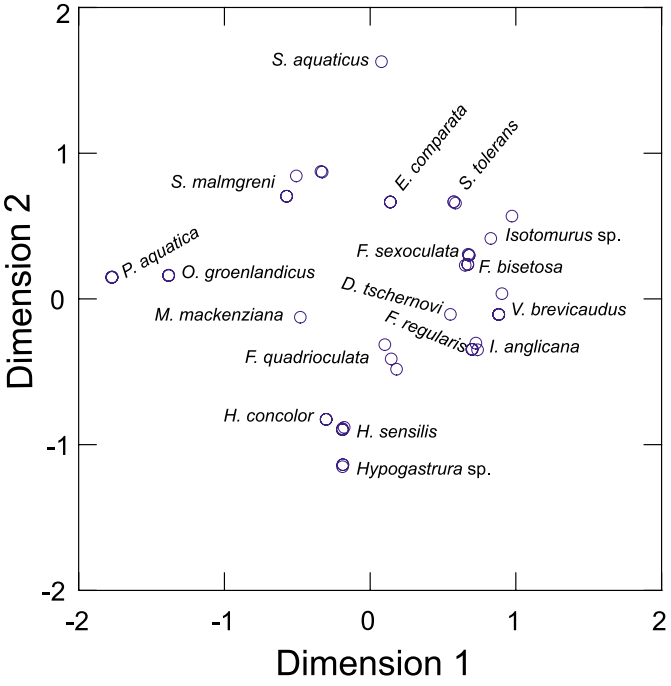
The neighbour joining analysis showed that species in families represented by more than one taxon (i.e., Isoto-

Table 2. Mean percentage divergence values (range given in parentheses, where applicable) between collembolan taxa collected from the Canadian Arctic.

	Individuals within species		Species within genus		Species between genera		Species between families	
Family	*n*	Divergence	*n*	Divergence	*n*	Divergence	*n*	Divergence
Neanuridae							1	23 (20–25)
Entomobryidae	1	0.00					1	21 (18–27)
Hypogastruridae	3	0.03 (0–0.05)	1	19 (8–25)			3	21 (16–25)
Isotomidae	9	0.97 (0–13.21)	1	17 (16–19)	7	19 (16–23)	10	22 (18–27)
Onychiuridae	1	0.00					1	22 (21–25)
Poduridae	1	0.00					1	25 (23–28)
Sminthuridae	1	3.68 (0.16–5.52)	1	21			2	23 (20–28)

Note: For individuals within species, *n* = number of species that had two or more individuals; for species within genus, *n* = number of genera with two or more species; for species between genera, *n* = number of genera analysed within a family; and for species between families, *n* = total number of species analysed.

Fig. 2. Multidimensional scaling of Euclidian distances for cytochrome-*c* oxidase subunit I (COI) gene sequences from 19 species (54 individuals) of Collembola collected from the Canadian Arctic. Circles identify individuals of each species included in the profile. In each case, individuals of the same nominate species grouped together.



midae, Hypogastruridae, and Sminthuridae) usually formed cohesive assemblages (Fig. 3). For the Isotomidae, which had the greatest species diversity, some association with other genera (e.g., *Sminthurides*) was observed, and *A. polaris* was located externally to other members of this family (Fig. 3).

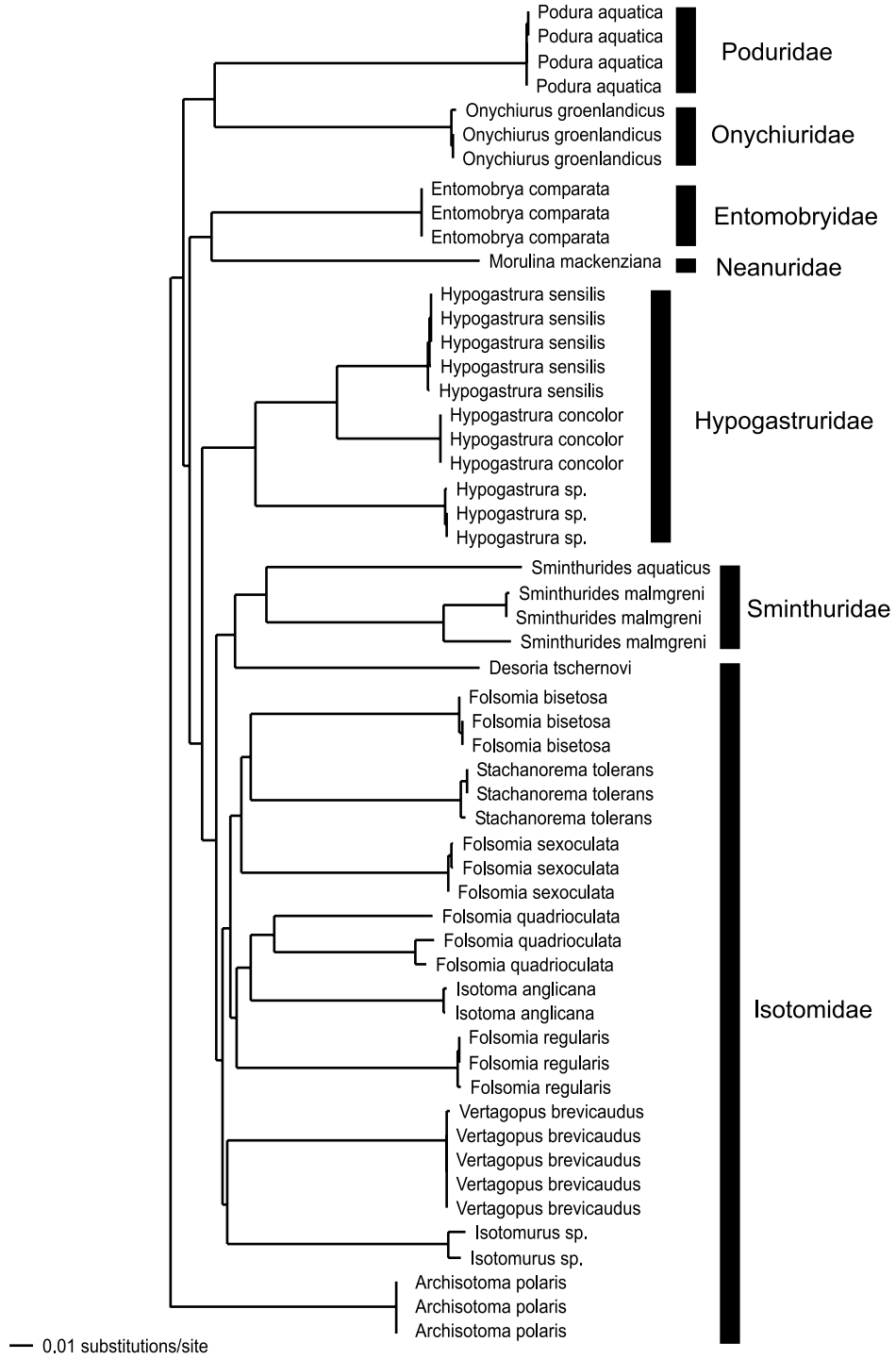
Discussion

In all cases we were able to successfully discriminate known species of Arctic Collembola using COI sequences. The A–T nucleotide bias commonly observed in hexapods was also detected in this study. However, the level of 60% observed in this study was lower than the level of 63%–64% reported for Antarctic Collembola (Frati et al. 1997; Stevens and Hogg 2003) and considerably less than the level of 70%–75% reported for some insect taxa (Nardi et al. 2003). In most cases, levels of divergence within a species were <1%, whereas divergences among species always exceeded 8%. Similar levels of divergence have been reported in other studies of springtails (Carapelli et al. 1995*a*). However, Frati et al. (2000*a*) reported divergences of up to 6% in Italian species of *Orchesella*. This suggests that sequence divergences greater than 8% are the rule among recognized species of Collembola, even when those species belong to a single genus. High intraspecific divergences were detected in two of the taxa examined in this study (5% for *S. malmgeni* and 13% for *F. quadrioculata*), but we expect that further study will reveal that these are cases in which currently undescribed sibling species were analyzed, particularly in the case of *Folsomia*. Morphologically cryptic species are well known among the Collembola (Carapelli et al. 1995*b*; Stevens and Hogg 2003) and among other hexapods (e.g., Jackson and Resh 1992; Hogg et al. 2002).

Levels of divergence between congeneric species of isotomid Collembola (mean = 19%; Table 2) were substantially higher than those reported in lepidopterans (mean = 6%) (Hebert et al. 2003*a*). In fact, collembolan divergences appear higher than those reported for animals at large: congeneric taxa showed an average of 11.2% sequence divergence (Hebert et al. 2003*b*). Our data support previous work (Carapelli et al. 1995*a*; Stevens and Hogg 2003) showing that the sequencing of mitochondrial genes is useful for discriminating closely related species of Collembola. For example, Carapelli et al. (1995*a*) were able to discriminate between closely related species of *Isotomurus* using the cytochrome-*c* oxidase subunit II gene, whereas Stevens and Hogg (2003) utilized the COI gene to discriminate between two sympatric lineages of the Antarctic springtail *Gomphiocephalus hodgsoni*. In the latter study, COI divergences were <2%, but allozyme data suggested that these differences reflected reproductively isolated species. These collective results demonstrate the effectiveness of mtDNA and, in particular, COI sequences for detecting species-level differences even in cases of recent divergence.

We conclude that COI barcodes will provide a powerful tool for species discrimination in the Collembola. Our work

Fig. 3. Neighbour joining analysis of uncorrected *p* distances based on an analysis of a 634-bp fragment of the COI gene.



has begun assembly of the COI database needed to create an identification system for Arctic Collembola, and it has also provided tentative evidence of the need for taxonomic revision, as indicated by our detection of putative new species in two genera. Further work should aim to extend sample sizes, taxonomic diversity, and geographic coverage. Such work will enable the creation of a DNA-based identification system for Collembola, enabling their identification to be executed by anyone with access to a basic DNA sequencing laboratory.

Acknowledgments

Penny Greenslade aided our initial identification work, but we are particularly grateful to Arne Fjellberg for his identification or confirmation of our identifications. Mark Stevens

and two anonymous reviewers provided valuable comments on earlier versions of the manuscript, and Shelley Ball, Sujeevan Ratnasingham, Jonathan Witt, Louise Hathaway, and Mark Stevens provided advice on the DNA analyses. Funding was provided through an NSERC operating grant to P.D.N.H. and through support from the University of Waikato and Antarctica New Zealand to I.D.H. We also thank the Polar Continental Shelf Project and the Nunavut Science Institute for logistical support in the Arctic.

References

Babenko, A. 1994. *Stachanorema tolerans* sp. nov. (Collembola; Isotomidae) from the Canadian High Arctic. Pol. Pismo Entomol. **63**: 23–27.

Babenko, A.B. 1993. Collembola in polar desert landscapes of the Devon Island, Canada N.W.T. Zool. J. **8**: 28–35. [In Russian.]

Carapelli, A., Fanciulli, P.P., Frati, F., and Dallai, R. 1995*a*. The use of genetic markers for the diagnosis of sibling species in the genus *Isotomurus* (Insecta, Collembola). Boll. Zool. **62**: 71–76.

Carapelli, A. Frati, F., Fanciulli, P.P., and Dallai, R. 1995*b*. Genetic differentiation of six sympatric species of *Isotomurus* (Collembola, Isotomoidae): is there any difference in their microhabitat preference? Eur. J. Soil Biol. **31**: 87–99.

Christiansen, K., and Bellinger, P. 1980. The Collembola of North America north of the Rio Grande. Grinnell College, Grinnell, Iowa.

Danks, H.V. 1981. Arctic arthropods: a review of systematics and ecology with particular reference to the North American fauna. Entomological Society of Canada, Ottawa, Ont.

Daugherty, C.H., Cree, A., Hay, J.M., and Thompson, M.B. 1990. Neglected taxonomy and continuing extinctions of tuatara (*Sphenodon*). Nature (Lond.), **347**: 177–179.

Fjellberg, A. 1986. Collembola of the Canadian high arctic. Review and additional records. Can. J. Zool. **64**: 2386–2390.

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

Folsom, J.W. 1937. Nearctic Collembola or springtails, of the family Isotomidae. U.S. Natl. Mus. Bull. No. 168. pp. 1–144.

Frati, F., Simon, C., Sullivan, J., and Swofford, D.L. 1997. Evolution of the mitochondrial cytochrome oxidase II gene in Collembola. J. Mol. Evol. **44**: 145–158.

Frati, F., Dell'Ampio, E., Casasanta, S., Carapelli, A., and Fanciulli, P. 2000*a*. Large amounts of genetic divergence among Italian species of the genus *Orchesella* (Insecta, Collembola) and the relationships of two new species. Mol. Phylogenet. Evol. **17**: 456–461.

Frati, F., Fanciulli, P.P., Carapelli, A., Dell'ampio, E., Nardi, F., Spinsanti, G., and Dallai, R. 2000*b*. DNA sequence analysis to study the evolution of Antarctic Collembola. Ital. J. Zool. Suppl. 1. pp. 133–139.

Hebert, P.D.N., Cywinska, A., Ball, S.L., and deWaard, J.R. 2003*a*. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B Biol. Sci. **270**: 313–322.

Hebert, P.D.N., Ratsingham, S., and deWaard, J.R. 2003*b*. Barcoding animal life: COI divergences among closely related species. Proc. R. Soc. Lond. B Biol. Sci. **270**(Suppl. 1): S96–S99.

Hogg, I.D., and Stevens, M.I. 2002. Soil fauna of Antarctic coastal landscapes. Ecol. Stud. **154**: 265–280.

Hogg, I.D., Willmann-Huerner, P., and Stevens, M.I. 2002. Population genetic structures of two New Zealand stream insects: *Archichauliodes diversus* (Megaloptera) and *Coloburiscus humeralis* (Ephemeroptera). N.Z. J. Mar. Freshw. Res. **36**: 491–501.

Jackson, J.K., and Resh, V.H. 1992. Variation in genetic structure among populations of the caddisfly *Helicopsyche borealis* from three streams in northern California, U.S.A. Freshw. Biol. **27**: 29–42.

Lessa, P. 1990. Multidimensional scaling of geographic genetic structure. Syst. Zool. **39**: 242–252.

Lipscomb, D., Platnick, N., and Wheeler, Q. 2003. The intellectual content of taxonomy: a comment on DNA taxonomy. Trends Ecol. Evol. **18**: 65–66.

Nardi, F., Spinsanti, G., Boore, J.L., Carapelli, A., Dallai, R., and Frati, F. 2003. Hexapod origins: monophyletic or paraphyletic? Science (Wash., D.C.), **299**: 1887–1889.

Petersen, H., and Luxton, M. 1982. A comparative analysis of soil fauna populations and their role in decomposition processes. Oikos, **39**: 287–388.

Savage, J.M. 1995. Systematics and the biodiversity crisis. Bioscience, **45**: 673–679.

Stevens, M.I., and Hogg, I.D. 2003. Long-term isolation and recent range expansion revealed for the endemic springtail *Gomphiocephalus hodgsoni* from southern Victoria Land, Antarctica. Mol. Ecol. **12**: 2357–2369.

Swofford, D.L. 2002. PAUP*: Phylogenetic Analysis Using Parsimony. Version 4.0*b*10 for Macintosh [computer program]. Sinauer Associates Inc., Sunderland, Mass.

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H., and Vogler, A.P. 2003. A plea for DNA taxonomy. Trends Ecol. Evol. **18**: 70–74.