CSIRO PUBLISHING

Invertebrate Systematics, 2012, **26**, 470–477
http://dx.doi.org/10.1071/IS12026

Challenging species delimitation in Collembola: cryptic diversity among common springtails unveiled by DNA barcoding

D. Porco[A,H], *A. Bedos*[B], *Penelope Greenslade*[C], *C. Janion*[D], *D. Skarżyński*[E], *M. I. Stevens*[F], *B. Jansen van Vuuren*[G] *and L. Deharveng*[B]

[A]Université de Rouen – Laboratoire ECODIV, Bâtiment IRESE A, Place Emile Blondel, 76821 Mont Saint Aignan Cedex, France.
[B]Muséum National d'Histoire Naturelle, UMR7205 «Origine, Structure et Evolution de la Biodiversité», 45 rue Buffon, CP50, 75005 Paris, France.
[C]Environmental Management, School of Science, Information Technology and Engineering, University of Ballarat, Ballarat, Vic. 3353, Australia.
[D]Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Private Bag x1, Matieland 7602, South Africa.
[E]Zoological Institute, Wrocław University, Przybyszewskiego 63/77, 51-148 Wrocław, Poland.
[F]South Australian Museum, GPO Box 234, Adelaide, SA 5000, and School of Earth and Environmental Sciences, University of Adelaide, SA 5005, Australia.
[G]Centre for Invasion Biology, Department of Zoology, University of Johannesburg, PO Box 524, Auckland Park 2006, South Africa.
[H]Corresponding author: Email: david.porco.gm@gmail.com

Abstract. Collembola is one of the major functional groups in soil as well as a model taxon in numerous disciplines. Therefore the accurate identification of specimens is critical, but could be jeopardised by cases of cryptic diversity. Several populations of six well characterised species of springtails were sequenced using the COI barcode fragment as a contribution to the global Collembola barcoding campaign. Each species showed high intraspecific divergence, comparable to interspecific sequence divergence values observed in previous studies and in 10 congeneric species barcoded here as a reference. The nuclear marker, 28S, confirmed all the intraspecific lineages found with COI, supporting the potential specific status of these entities. The implications of this finding for taxonomy and for disciplines relying on species names, such as evolution and ecology, are discussed.

Received 15 April 2012, accepted 22 September 2012, published online 19 December 2012

Introduction

The widespread occurrence of Collembola in a wide range of habitats allied to their critical ecological role as decomposers, both directly by consumption of decaying vegetal material and indirectly by influencing the distribution and growth of fungal species, makes this hexapod class one of the most important functional groups in soil (Hopkin 1997). With ~8000 described species, taxon richness in the class appears relatively low in comparison with other soil taxa such as mites (Hammond 1992). However, this estimate represents only a fraction of the actual number of species present on the planet (Deharveng 2004). Even in the most thoroughly studied regions of the world, such as Western Europe, where most species are thought to be known, new collembolan species are still being discovered (Kovac and Papac 2010; Busmachiu *et al*. 2010; Lukic *et al*. 2010; Porco *et al*. 2010*a*; Jordana *et al*. 2011*a*, 2011*b*). The number of new species is undoubtedly much higher in lesser known regions of the world, particularly the tropics (Cassagnau 1991; Palacios-Vargas and Arbea 2009). As is often stressed, taxonomic expertise is the main bottleneck for progress in knowledge of Collembola, as well as limiting species information that is available to the wider scientific community (Stevens *et al*. 2011). Molecular taxonomy, and in particular DNA barcoding (Hebert *et al*. 2003), which has been rapidly developing in the last 10 years, is part of the solution to overcome this gap, by providing access for the non-specialist to a powerful tool both for species identification and for the detection of new taxa (Stevens *et al*. 2011).

One of the most important difficulties in assessing collembolan species richness is the high level of cryptic diversity suspected in many species, including several in Europe that are cosmopolitan in distribution (Deharveng 2004). Karyotypic or genetic intraspecific divergences suggesting the occurrence of cryptic diversity has already been identified in various species of Collembola (Cassagnau 1976; Fanciulli *et al*. 1991, 2000, 2009; Frati *et al*. 1992; Carapelli *et al*. 1995*a*; Stevens *et al*. 2006; Cicconardi *et al*. 2010; Torricelli *et al*. 2010).

Journal compilation © CSIRO 2012

We approach the potential problem of cryptic diversity here using the DNA barcoding fragment and the nuclear marker 28S (D2 region) of six common species from four Collembola families, currently considered to be morphologically well defined.

The investigation of the molecular diversity of these six well known species will serve to evaluate the potential contribution of molecular taxonomy tools to the Collembola.

Materials and methods

Sampling

Six species (*Bilobella aurantiaca* (Caroli, 1912), *Deutonura monticola* (Cassagnau, 1954), *Ceratophysella denticulata* (Bagnall, 1941), *Heteromurus major* (Moniez, 1889), *Tomocerus baudoti* Denis, 1932 and *Podura aquatica* Linnæus, 1758) were sampled from various locations worldwide as a part of the global Collembola DNA barcoding campaign along with 10 congeneric species (*Bilobella massoudi* Cassagnau, 1968, *Ceratophysella cavicola* (Börner, 1901), *Ceratophysella pseudarmata* (Folsom, 1916), *Ceratophysella granulata* Stach, 1949, *Deutonura caerulescens* Deharveng, 1982, *Deutonura gibbosa* Porco, Bedos & Deharveng, 2010, *Deutonura provincialis* Deharveng, 1979, *Heteromurus nitidus* (Templeton, 1836), *Tomocerus minor* (Lubbock, 1862), *Tomocerus problematicus* Cassagnau, 1964) and barcoded to provide baseline data for species divergences within genera (Table 1). Specimens were collected using Berlese funnels and preserved in 96% ethanol. A total of 204 specimens belonging to the 16 species were sequenced.

Molecular analysis

DNA was extracted from entire specimens in 30 µL of lysis buffer 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 further morphological examination (Porco *et al*. 2010*b*). 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). A standard PCR reaction protocol was used for amplifications, 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 ver. 3.1 on an ABI 3730 DNA Analyzer (Applied Biosystems). Sequences were assembled with Sequencer 4.5 (GeneCode Corporation, Ann Arbor, MI, USA) and aligned by eye using BIOEDIT ver. 7.0.5.3 (Hall 1999).

In addition, for 83 of the sampled specimens, a fragment of the D2 region of the nuclear gene 28S of 444 base pairs was amplified and sequenced under the same conditions as COI. The primer pair used was 'D2coll' and 'C2coll', designed specifically for Collembola (D'Haese 2002). As we observed no indels in the COI and 28S sequences, sequence alignment was unambiguous. Sequences are publicly available on BOLD (project DATASET-CRYCOL1) and on GenBank (Supplementary material Table 1).

Data analysis

Distance analyses were performed with MEGA4 software (Tamura *et al*. 2007). Neighbour-joining (Saitou and Nei 1987) algorithm with the Kimura-2 parameter model (Kimura 1980) was used to estimate the genetic distances. The robustness of nodes was evaluated through bootstrap reanalysis of 1000 pseudoreplicates. The trees were replotted using the online utility iTOL (Letunic and Bork 2007).

Results

Deep genetic divergences were found for COI both between and within species (Fig. 1). For the 10 species selected as a standard for each genus examined here, the overall mean interspecific divergence was 27.66% (ranging from 21.09% to 35.09%) and the mean intraspecific divergence was 0.52% (ranging from 0% to 1.4%) (Supplementary material Table 1 and Table 2). Among these 10, the mean interspecific divergence between congeneric species was 22.58% (Supplementary material Table 1).

By contrast, the six targeted species displayed high intraspecific variation for COI, from 11.33% to 21.47%, which is of the same order of magnitude found between morphologically defined species of the genus, with an exception being *B. aurantiaca* (intralineage divergence 11.33%, interspecific 23.19%) (Table 2). These species included 2–7 COI lineages (Table 2), and some were found to be sympatric (Table 1). All these lineages were confirmed by 28S sequences with a mean intralineage divergence ranging from 0% to 0.25% and mean interlineages from 0.61 to 9.53 (Table 2). The lowest values were found between the lineages of *B. aurantiaca*.

Discussion

Specific diversity of the group

These preliminary results of the Collembola barcoding campaign illustrate a high degree of cryptic diversity within common species that previously were considered in the taxonomic literature to be well characterised single taxa. High intraspecific values of divergence support specific status for each of these lineages, as they are consistent with interspecific divergences observed so far in Collembola ranging from 8% to 22% (Hogg and Hebert 2004; Stevens and Hogg 2006; Rougerie *et al*. 2009; McGaughran *et al*. 2010; Porco *et al*. 2010*a*, 2012*a*). Moreover, these intraspecific divergences were confirmed by the variations found in a slowly evolving nuclear marker 28S (Table 2). These results support the hypothesis of reproductive isolation among populations of these lineages for a substantial period. The age between interspecific and interlineage divergence (Table 2) was estimated on the basis of the COI models applied by Stevens *et al*. (2006) using an arthropod strict molecular clock COI calibration of 1.5–2.3% sequence divergence per million years. We found a minimum interspecific age of 8.52–13.07 million years (within *Tomocerus*) and maximum of 10.1–17.88 million years (within *Deutonura*). The minimum interlineage age of 4.93–7.55 million years was found within *B. aurantiaca* and a maximum of 9.33–14.31 million years between *H. major* lineages. The 28S divergence between

Table 1. Populations sampled in the species targeted
Lineages found, type of microhabitat and collection dates are shown. NA, not available

Species	Country	Locality	*N*	Lineages	Habitat/subtract	Collection date
Bilobella aurantiaca	France	Ariège (Aulus-les-Bains)	6	L1	Beech forest, decaying wood	13 Aug. 2004
	France	Pyrénées-Orientales (Corsavy)	7	L2	Broad-leaved bush	27 Jun. 2005
	France	Pyrénées-Orientales (Mosset)	6	L3	Meadow, decaying pine trunk	02 Jun. 2010
	France	Ariège (Orlu)	3	L3	Fir-beech forest, decaying wood	06 Sep. 2001
Bilobella massoudi	Slovenia	Planina	5		Spruce-beech litter	01 Sep. 2010
	Slovenia	Postojna	1		Beech litter	01 Sep. 2010
	Slovenia	Goricane	2		Beech forest, decaying wood	01 Sep. 2010
Ceratophysella cavicola	Poland	Sudetes (Karkonosze Mountains)	4		Litter in adit	06 Jun. 2009
Ceratophysella denticulata	France	Pyrénées-Orientales (Corsavy)	4	L1	Beech forest, decaying wood and litter	27 Jun. 2005
	United Kingdom	South East (London)	2	L2	Beech forest, litter	24 Jun. 2003
	South Africa	Western Cape (Cape Town, Kirstenbosch)	5	L3	Forest litter	26 Aug. 2011
	Australia	Macquarie Island (Buckles Bay)	3	L3	Tussock grass, litter	08 Jan. 2004
	Canada	Manitoba (Churchill)	5	L3	Pond bank, moss	21 Aug. 2006
	New Zealand	Mt Stokes Track	6	L3	Forest litter	11 Dec. 2007
	New Zealand	South Island (Southland Catlins)	1	L3	Forest litter	01 Apr. 2007
Ceratophysella granulata	Poland	Carpathians (Tatra Mountains)	5		Dwarf pine shrubs, litter and mosses	14 Aug. 2009
Ceratophysella pseudarmata	Canada	Manitoba (Churchill)	6		Shrubs, decaying wood	11 Aug. 2006 27 Jul. 2009
	Canada	Ontario (Guelph)	2		Forest litter	14 May 2009
Deutonura caerulescens	France	Var (Toulon)	6		Open garrigue, litter	17 Dec. 2008
Deutonura gibbosa	France	Isère (Saint-Quentin-sur-Isère)	2		Mixed broad-leaved forest litter	03 Nov. 2007
	France	Hautes-Alpes	3		Beech forest litter	04 Nov. 2007
Deutonura monticola	France	Haute-Garonne (Antichan-de-Frontignes)	2	L1	Beech forest, decaying wood and litter	18 May 2006
	France	Haute-Garonne (Juzet-d'Izaut)	4	L1	Beech forest, decaying wood and litter	06 Jun. 2006
	France	Ariège (Le Port)	2	L1, L2	Beech forest litter	14 Aug. 2008
	France	Ariège (Orlu)	6	L2, L3	Fir-beech forest, decaying wood	07 Jun. 2005
	France	Aude (Counozouls)	5	L3	Beech-fir forest litter	02 Jun. 2010
	France	Pyrénées (Caudies-de-Fenouilledes)	2	L3	Maple litter	03 May 2007
	France	Pyrénées-Orientales (Corsavy)	2	L3	Beech forest litter	27 Jun. 2005
Deutonura provincialis	France	Alpes-Maritimes (Peille)	5		Holm-oak forest litter	15 Nov. 2008
Heteromurus major	France	Pyrénées-Orientales (Corsavy)	6	L1, L5	Maple and oak forest litter	27 Jun. 2005
	Belgium	West-Vlaanderen (Bruges)	2	L2	Broad-leaved wood litter (city park)	12 Jun. 2004
	France	Corsica (Bonifacio)	5	L3	Holm-oak forest litter	19 Mar. 2008
	France	Pyrénées-Orientales (Caudies)	6	L4	Holm-oak forest litter	14 Feb. 2005
	France	Pyrénées-Orientales (Prats-de-Mollo-La Preste)	5	L5	Beech forest litter	14 Sep. 2004
	France	Alpes-Maritimes (Peille)	6	L6	Pine and maple forest litter	15 Nov. 2008
	Spain	Navarra (Aritzkuren)	5	L7	Oak forest litter	10 Aug. 2008
Heteromurus nitidus	France	Essone (Brunoy)	4		Laboratory strains	
	France	Haute-Garonne (Juzet)	4		Laboratory strains	
Podura aquatica	Canada	Nunavut (Bylot Island)	4	L1	NA	25 Jun. 2007

(*continued next page*)

Table 1. (*continued*)

Species	Country	Locality	*N*	Lineages	Habitat/subtract	Collection date
	Canada	Nunavut (Qikiqtaaluk, Western Bylot Island)	3	L1	NA	21 Jun. 2008
	Canada	Manitoba (Riding Mountain Nat. Park)	5	L2	Wet meadow/marsh	15 Aug. 2008
	Canada	Alberta (Elk Island Nat. Park)	5	L2	Deciduous forest pond	10 Aug. 2010
	Canada	Manitoba (Churchill)	9	L2	Pond bank	10 Jul. 2008
Tomocerus baudoti	France	Alpes-Maritimes (Peille)	6	L1	Maple and pine forest litter	15 Nov. 2008
	France	Dordogne (Villac)	4	L2	Hornbeam and chestnut forest litter	08 Sep. 2004
Tomocerus minor	France	Pyrénées-Orientales (Prats-de-Mollo-La Preste)	6		Beech forest litter	15 Sep. 2008
Tomocerus problematicus	France	Pyrénées (Lacourt)	3		Cave	10 Oct. 2004
	France	Pyrénées (Merigon)	4		Cave	11 Feb. 2003

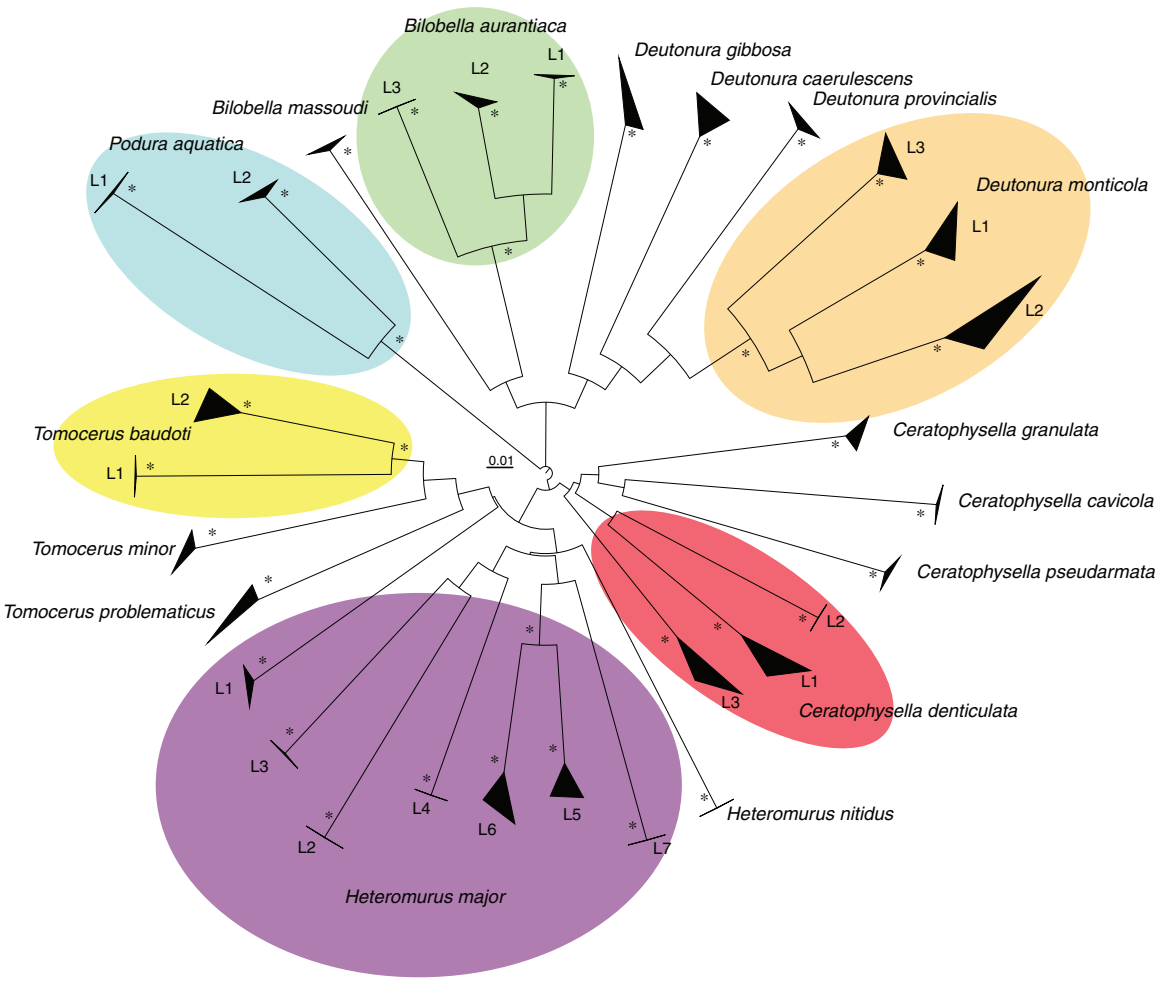

Fig. 1. Neighbour-joining trees (K2P) for the 30 species and lineages based on the COI. Bootstrap support values ≥99% are shown on branches with an asterisk. The upper and lower sides of each triangle represent respectively the maximum and minimum genetic distances within a species.

lineages of *B. aurantiaca* was the lowest found in the species and lineages investigated here, and may be the result of later divergence within this species.

Including these species-level lineages would increase the total species count of these taxa from 16 to 30. Given the limited number of populations examined for the targeted species

Table 2. Inter/intraspecific and intralineages genetic divergences for COI and 28S
(*calculated from the pairwise comparison among species and lineages (not between lineages))

Species	No. of lineages	Mean Intraspecific K2P divergences COI	Mean Intralineages K2P divergences COI	Mean Interlineages K2P divergences COI	Mean Interspecific K2P divergences COI[A]	Mean Intralineages K2P divergences 28S	Mean Interlineages K2P divergences 28S
Bilobella aurantiaca	3		0.24	11.33	23.19	0.00	0.61
Bilobella massoudi	1	0.26			23.19		
Ceratophysella cavicola	1	0.23			22.66		
Ceratophysella denticulata	3		1.95	19.51	22.66	0.25	7.88
Ceratophysella granulata	1	1.08			22.66		
Ceratophysella pseudarmata	1	0.16			22.66		
Deutonura caerulescens	1	1.40			23.24		
Deutonura gibbosa	1	0.13			23.24		
Deutonura monticola	3		1.62	17.17	23.24	0.09	1.33
Deutonura provincialis	1	0.49			23.24		
Heteromurus major	7		0.35	21.47	23.02	0.00	2.66
Heteromurus nitidus	1	0.00			23.02		
Podura aquatica	2		0.30	19.21		0.00	4.41
Tomocerus baudoti	2		0.15	21.17	19.60	0.00	9.53
Tomocerus minor	1	1.27			19.60		
Tomocerus problematicus	1	0.21			19.60		

[A]Calculated from the pairwise comparison among species and lineages (not between lineages).

compared with their total distribution, the number of COI lineages is expected to increase with broader sampling. The mismatch between morphological species and molecular species-level entities is considerable, suggesting that global species richness of the class Collembola is much higher than currently estimated (Cicconardi *et al*. 2010). Even if not recognised and named as cryptic diversity, a high level of genetic variation or polytypism has already been demonstrated within other species of Collembola. The first evidence was provided by detailed analyses of polytene chromosomes in the salivary glands of Neanurinae: in the two species examined (*Bilobella aurantiaca* and *Lathriopyga longiseta*), extreme levels of polytypism were detected (Cassagnau 1976; Cassagnau *et al*. 1979). In our study, three lineages of *B. aurantiaca* were detected. Other common European species such as *Tetrodontophora bielanensis* Reuter, 1882 (Fanciulli *et al*. 1991), *Orchesella villosa* Linnaeus, 1767 (Frati *et al*. 1992), *Tomocerus vulgaris* (Tullberg, 1871) (Fanciulli *et al*. 2000), *Isotomurus palustris* (Müller, 1776) (Carapelli *et al*. 1995*a*), and *Allacma gallica* (Carl, 1899) (Fanciulli *et al*. 2009) also have been shown to exhibit high intraspecific genetic variation. Recently, cryptic diversity was documented in several species of *Lepidocyrtus* (Cicconardi *et al*. 2010), in the Antarctic species *Friesea grisea* (Schäffer, 1891) (Torricelli *et al*. 2010) and in *Cryptopygus antarcticus* Willem, 1901 (Stevens *et al*. 2006; McGaughran *et al*. 2010). Thus, in various species of Collembola, genetic differentiation does not conform with morphological boundaries among species. In other cases, morphological differentiation may exist but has been overlooked. Colour patterns, long considered as polymorphic in several groups of Collembola, have been found to be a valid character to discriminate between species when molecular characters were used. The emblematic cases in Collembola are those of the *Isotomurus palustris* (Carapelli *et al*. 1995*b*; Frati *et al*. 1995) and *Lepidocyrtus dispar* Mari Mutt, 1986/*biphasis* Mari Mutt, 1986 complexes (Soto-Adames 2002).

As DNA barcoding discloses high level of cryptic diversity, it offers an additional contribution to collembolan biodiversity, showing that traditional approaches were unable to detect a significant part of it. We show that collembolan diversity can be considered at two levels: the traditional, coarser, morphological level currently used by most taxonomists and ecologists, and a higher-resolution level using molecular approaches.

Taxonomy

The discovery of these cases of cryptic diversity also provides an opportunity to explore unexpected species complexes (Stevens *et al*. 2011). The current study detected several molecular lineages that can be designated as MOTUs (Molecular Operational Taxonomic Units) while awaiting further morphological characterisation. Meanwhile, these MOTUs should be named so they can be referred to in other studies. This could be achieved by the morphological examination of voucher specimens after DNA extraction (Porco *et al*. 2010*b*), in particular by reconsidering characters that were first rejected for diagnosis (Carapelli et al. 1995*a*; Soto-Adames 2002) or even by examining those morphological characters used in the past to assign synonyms. So far, we have not detected such characters in our material. In such a case, the alternative solution is to use molecular taxonomy to match the MOTUs with type specimens in order to designate which one should bear the species name. A first option would be to sequence mini-barcodes of the holotype or paratype specimens, as previously reported for moths (Hausmann *et al*. 2009) and earthworms (James *et al*. 2010). However, this is likely

to be unsuccessful because of the slide mounting that Collembola specimens undergo for identification, as well as DNA decay in poor preservation conditions over a long period. Sequencing of mini-barcodes of unmounted specimens from the type series could be an alternative solution. In order to avoid dealing with ancient DNA problems when further cryptic lineages are uncovered, we recommend sequencing the barcode region of fresh unmounted specimens of the type series whenever a new species is described (Porco *et al*. 2010*a*; Palacios-Vargas *et al*. 2011).

A second option is to barcode fresh specimens collected from the type locality, as it was done in *Parisotoma notabilis* (Schäffer, 1896) (Porco *et al*. 2012*a*). This allows linking of the topotype population, from which the type specimen was likely described, to one of the detected MOTUs (which could be named 'MOTU-type'). The results should only be considered as a working hypothesis as (1) the MOTU sequenced may not be the one from which the species was described because of changes in distribution or (2) several MOTUs could occur in the same place.

Biological and ecological implications

Some of the species targeted here have been used in ecological (Salmon 2001; Rocha-Ramírez *et al*. 2007), physiological (Rost 2006) and ecotoxicological (Lupetti *et al*. 1994) studies. As several lineages are present within these species, some of the results of these studies may need reexamination, especially as high genetic divergences, underlying a potential specific status, may be associated with different ecologies. An example is shown by the six species of *Isotomurus* detected within the *I. palustris* complex. Each species had different microhabitat preferences that were related to different ecological traits (Carapelli *et al*. 1995*b*). Sampling in the present study was not designed to examine possible different population ecologies. Even so, the two MOTUs found for *P. aquatica* have different distributions, one in the temperate zone and the other in the Arctic, suggesting probable different ecological traits.

These ecological differences between cryptic species highlight possible errors that could be made in bioindication, ecotoxicological or invasion biology studies if several distinct genetic entities are considered to be a single taxon. For instance, *Folsomia candida* Willem, 1902 has become a standard species in the assessment of environmental risks (Riepert and Kula 1996; Crouau *et al*. 2009), but substantial genetic divergences have been found among strains currently used in the laboratory and in wild populations of this species (Chenon *et al*. 2000; Frati *et al*. 2004; Tully *et al*. 2006). The links between intraspecific diversity and differences in characteristics for its application in bioindication were established by Barateiro Diogo *et al*. (2007). Nevertheless the strains diverged by only 1.39% for a COI fragment. By contrast, Frati *et al*. (2004) found high divergence among wild populations of this species for another mitochondrial gene COII (16–18%). In this latter case, much more variation in biological and ecological traits could be expected. The use of such genetically divergent strains could cause serious issue for comparing and interpreting results. However, nine European clones currently used in ecotoxicology have been tested and showed only minor differences in responses to toxic substances (Chenon *et al*. 2000), suggesting that the genetic divergence between these clones is much lower than that found by Frati *et al*. (2004) in wild populations.

In turn, ecological discrepancies in life traits, evident when compiling information from the literature (Potapov 2001), could suggest the presence of cryptic diversity that should be further investigated with molecular taxonomic tools. In this respect, species of wide ecological range and with eurytopic characteristics should be carefully scrutinised. Another critical issue would result from studies based on 'chimeric species' composed of mixed populations of sympatric lineages, such as some of the ones found here (*D. monticola* L2 and L3 and *H. major* L1 and L5) (Table 1). Thus, COI sequencing has to be recommended in species where such cryptic diversity is detected, in order to verify on which cluster the results are based.

Species distribution

Species with a wide distribution could include several MOTUs, each with a more restricted distribution. Some of the species sampled here have a limited distribution while others have a broader intercontinental one. Cryptic lineages could alter perceptions of invasive collembolan species, as not all the lineages, or perhaps only one, may be invasive (Porco *et al*. 2012*b*). Endemic taxa could therefore be unrecognised, with consequences for regional conservation policies. Several cases illustrate this concern in our dataset. The cosmopolitan species *C. denticulata* appears, on the limited sample at hand for this study, to be an assemblage of at least three well differentiated MOTUs found in three different regions, although all individuals found outside Europe belong to a single lineage. The species *B. aurantiaca* and *D. monticola* have western Mediterranean and south-western European distribution ranges respectively, but our results showed that they comprise a group of MOTUs, each with a more restricted distribution. The sampling of only five Canadian populations of *P. aquatica* revealed two highly divergent entities found in two distinct areas (arctic and subarctic/temperate). The case of *H. major*, which is distributed throughout Europe, is even more extreme, as each of its seven recognised MOTUs are known only from a single population. Geographic structure seems to be normal in species of different families. This suggests that widely distributed species, which are also dominant in many communities, should be investigated as a priority, as further sampling across their whole distribution is likely to disclose several MOTUs illustrating fine-grain biogeographical patterns, which will contribute to elucidating the phylogeography of the species.

Phylogeny

In this study, sequences from both a mitochondrial and a nuclear marker have been found to be significantly divergent between species-level MOTUs within morphologically well defined species. The phylogenetic position of a single one of these cryptic lineages will reflect only a part of the biological reality and may lead to conflicting hypotheses if studies use different lineages, as in some cases these lineages may not even belong to a single monophyletic group. Ignoring these sometimes numerous cryptic lineages (e.g. seven in *H. major*), could significantly bias

sampling for phylogenetic reconstructions of genera or small families, and so lead to incorrect hypotheses.

Conclusion

Our findings illustrate the value of systematic DNA barcoding campaigns for Collembola, and emphasise the value of the technique for biodiversity surveys that could otherwise considerably underestimate species diversity.

The large amount of cryptic diversity found suggests that many more cryptic lineages are yet to be uncovered. Considering all the potential issues generated by this situation, we suggest (1) to provisionally name the MOTUs discriminated within a species (i.e. numbering them as in this paper), so they can be recognised and used in different disciplines, and (2) to implement, whenever possible, the sequencing of COI to verify which lineages are involved. This will allow an accumulation of data on non-ambiguous species-level entities, for higher resolution and more accurate surveys.

Acknowledgements

This work was funded by NSERC, Genome Canada and the Ontario Genomics Institute (2008-OGI-ICI-03). We are grateful to Joël Bêty, Elise Bolduc, Martin Patenaude, Ludovic Jolicoeur and Laura McKinnon for the Bylot Island specimens of *P. aquatica* used in this study collected for the project ArcticWOLVES (Funding and logistic provided by grants from the Canadian federal government's International Polar Year program (Indian and Northern Affairs), the Natural Science and Engineering Research Council of Canada, the Canadian Ministry of Environment, and the Polar Continental Shelf Program (Natural Resources Canada)).

References

Barateiro Diogo, J., Natal-Da-Luz, T., Paulo Sousa, J., Vogt, C., and Nowak, C. (2007). Tolerance of genetically characterized *Folsomia candida* strains to phenmedipham exposure – a comparison between reproduction and avoidance tests. *Journal of Soils and Sediments* **7**, 388–392.

Busmachiu, G., Deharveng, L., and Weiner, W. M. (2010). A new species of the genus *Lathriopyga* Caroli, 1912 (Collembola, Neanuridae, Neanurinae) from the Republic of Moldova. *Zootaxa* **2639**, 53–58.

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). *Bollettino di Zoologia* **62**, 71–76. doi:10.1080/11250009509356053

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

Cassagnau, P. (1976). Variability of polytene chromosomes in *Bilobella aurantiaca* Caroli (Collembola, Neanuridae) its relations with biogeography and ecology of species. *Archives De Zoologie Experimentale Et Generale* **117**, 511–572. [In French]

Cassagnau, P. (1991). Les collemboles Neanurinae de l'Himalaya 2: Paranurini et Paleonurini paucitubercules. *Travaux du Laboratoire d'Ecobiologie des Arthropodes Edaphiques Toulouse* **6**, 1–20. [In French]

Cassagnau, P., Dallai, R., and Deharveng, L. (1979). Polytene chromosomes polymorphism in *Lathriopyga longiseta* Caroli (Collembola Neanuridae). *Caryologia* **32**, 461–483.

Chenon, P., Rousset, A., and Crouau, Y. (2000). Genetic polymorphism in nine clones of a parthenogenetic collembolan used in ecotoxicological testing. *Applied Soil Ecology* **14**, 103–110. doi:10.1016/S0929-1393(00)00047-0

Cicconardi, F., Nardi, F., Emerson, B. C., Frati, F., and Fanciulli, P. P. (2010). Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. *Molecular Ecology* **19**, 386–400. doi:10.1111/j.1365-294X.2009.04457.x

Crouau, Y., Cazes, L., and Bur, T. (2009). The use of the Collembola reproduction test in bioassays of polluted soils. In 'Progress in Environmental Science and Technology. Vol II, Pts A and B'. (Eds S. C. Li, Y. J. Wang, F. X. Cao, P. Huang and Y. Zhang.) pp. 1843–1845. (Science Press Beijing: Beijing.)

D'Haese, C. A. (2002). Were the first springtails semi-aquatic? A phylogenetic approach by means of 28S rDNA and optimization alignment. *Proceedings. Biological Sciences* **269**, 1143–1151. doi:10.1098/rspb.2002.1981

Deharveng, L. (2004). Recent advances in Collembola systematics. *Pedobiologia* **48**, 415–433. doi:10.1016/j.pedobi.2004.08.001

Fanciulli, P. P., Frati, F., Dallai, R., and Rusek, J. (1991). High genetic divergence among populations of *Tetrodontophora bielanensis* (Insecta, Collembola) in Europe. *Revue D Ecologie Et De Biologie Du Sol* **28**, 165–173.

Fanciulli, P. P., Melegari, D., Carapelli, A., Frati, F., and Dallai, R. (2000). Population structure, gene flow and evolutionary relationships in four species of the genera *Tomocerus* and *Pogonognathellus* (Collembola, Tomoceridae). *Biological Journal of the Linnean Society. Linnean Society of London* **70**, 221–238.

Fanciulli, P. P., Carapelli, A., Belloni, M., Dallai, R., and Frati, F. (2009). Allozyme variation in the springtails *Allacma fusca* and *A. gallica* (Collembola, Sminthuridae). *Pedobiologia* **52**, 309–324. doi:10.1016/j.pedobi.2008.11.004

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. *Molecular Marine Biology and Biotechnology* **3**, 294–299.

Frati, F., Fanciulli, P. P., and Dallai, R. (1992). Genetic diversity and taxonomy in soil-dwelling insects – the genus *Orchesella*. *The Journal of Heredity* **83**, 275–281.

Frati, F., Carapelli, A., Fanciulli, P. P., and Dallai, R. (1995). The genus *Isotomurus*: where molecular markers help to evaluate the importance of morphological characters for the diagnosis of species. *Polskie Pismo Entomologiczne* **64**, 41–51.

Frati, F., Negri, I., Fanciulli, P. P., Pellecchia, M., De Paola, V., Scali, V., and Dallai, R. (2004). High levels of genetic differentiation between *Wolbachia*-infected and non-infected populations of *Folsomia candida* (Collembola, Isotomidae). *Pedobiologia* **48**, 461–468. doi:10.1016/j.pedobi.2004.04.004

Hajibabaei, M., DeWaard, J. R., Ivanova, N. V., Ratnasingham, S., Dooh, R. T., Kirk, S. L., Mackie, P. M., and Hebert, P. D. N. (2005). Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B-Biological Sciences* **360**, 1959–1967. doi:10.1098/rstb.2005.1727

Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W., and Hebert, P. D. N. (2006). DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 968–971. doi:10.1073/pnas.0510466103

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95–98.

Hammond, P. M. (1992). Species inventory. In 'Global Biodiversity: Status of the Earth's Living Resources. A Report Compiled by the World Conservation Monitoring Centre'. (Ed. B. Groombridge.) pp. 17–39. (Chapman & Hall: London.)

Hausmann, A., Hebert, P. D. N., Mitchell, A., Rougerie, R., Sommerer, M., Edwards, T., and Young, C. J. (2009). Revision of the Australian *Oenochroma vinaria* Guenee, 1858 species-complex (Lepidoptera:

Geometridae, Oenochrominae): DNA barcoding reveals cryptic diversity and assesses status of type specimen without dissection. *Zootaxa* **2239**, 1–21.

Hebert, P. D. N., Cywinska, A., Ball, S. L., and DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **270**, 313–321. doi:10.1098/rspb.2002.2218

Hogg, I. D., and Hebert, P. D. N. (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. doi:10.1139/z04-041

Hopkin, S. P. (1997). 'Biology of the Springtails (Insecta: Collembola).' (Oxford University Press: New York & Tokyo.)

Ivanova, N. V., Dewaard, J. R., and Hebert, P. D. N. (2006). An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes* **6**, 998–1002. doi:10.1111/j.1471-8286.2006.01428.x

James, S. W., Porco, D., Decaëns, T., Richard, B., Rougerie, R., and Erséus, C. (2010). DNA barcoding reveals cryptic diversity in *Lumbricus terrestris* L., 1758 (Clitellata): resurrection of *L. herculeus* (Savigny, 1826). *PLoS ONE* **5**, e15629. doi:10.1371/journal.pone.0015629

Jordana, R., Schulz, H.-J., and Baquero, E. (2011*a*). New species of *Entomobrya* from Germany (Collembola, Entomobryini). *Soil Organisms* **83**, 249–264.

Jordana, R., Giuga, L., and Baquero, E. (2011*b*). New species of *Entomobrya* from Etna mountain, Sicily (Collembola Entomobryomorpha). *Redia-Giornale Di Zoologia* **94**, 35–38.

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. doi:10.1007/BF01731581

Kovac, L., and Papac, V. (2010). Revision of the genus *Neelus* Folsom, 1896 (Collembola, Neelida) with the description of two new troglobiotic species from Europe. *Zootaxa* **2663**, 36–52.

Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics (Oxford, England)* **23**, 127–128. doi:10.1093/bioinformatics/btl529

Lukic, M., Houssin, C., and Deharveng, L. (2010). A new relictual and highly troglomorphic species of Tomoceridae (Collembola) from a deep Croatian cave. *ZooKeys* **69**, 1–16.

Lupetti, P., Marsili, L., Focardi, S., and Dallai, R. (1994). Organochlorine compounds in litter-dwelling arthropods: Collembola (Insecta Apterygota) from central Italy. *Acta Zoologica Fennica* **195**, 94–97.

McGaughran, A., Torricelli, G., Carapelli, A., Frati, F., Stevens, M. I., Convey, P., and Hogg, I. D. (2010). Contrasting phylogeographical patterns for springtails reflect different evolutionary histories between the Antarctic Peninsula and continental Antarctica. *Journal of Biogeography* **37**, 103–119. doi:10.1111/j.1365-2699.2009.02178.x

Palacios-Vargas, J. G., and Arbea, J. I. (2009). Neotropical species of *Proisotoma* (Collembola: Isotomidae), with description of two new cave species from America. *Revista Mexicana De Biodiversidad* **80**, 445–453. [In Spanish]

Palacios-Vargas, J. G., Deharveng, L., and D'Haese, C. A. (2011). The genus *Pronura* (Collembola: Neanuridae) in South America, with descriptions of two new species and a barcode sequence for one of them. *Revue Suisse de Zoologie* **118**, 197–205.

Porco, D., Bedos, A., and Deharveng, L. (2010*a*). 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., Rougerie, R., Deharveng, L., and Hebert, P. (2010*b*). 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. doi:10.1111/j.1755-0998.2010.2839.x

Porco, D., Potapov, M., Bedos, A., Busmachiu, G., Weiner, W. M., Hamra-Kroua, S., and Deharveng, L. (2012*a*). Cryptic diversity in the ubiquist species *Parisotoma notabilis* (Collembola, Isotomidae): a long used chimeric species? *PLoS ONE* **7**(9), e46056.

Porco, D., Decaëns, T., Deharveng, L., James, S., Skarżyński, D., Erséus, C., Butt, K., Richard, B., and Hebert, P. N. (2012*b*). 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.* doi:10.1007/s10530-012-0338-2

Potapov, M. (2001). Synopses on Palaearctic Collembola. Volume 3. Isotomidae. *Abhandlungen und Berichte des Naturkundemuseums Goerlitz* **73**, 1–603.

Riepert, F., and Kula, C. (1996). Development of laboratory methods for testing effects of chemicals and pesticides on Collembola and earthworms. In 'Communications from the Federal Biological Institute for Agriculture and Forestry'. (Ed. Berlin-Dahlem.) pp. 82.

Rocha-Ramírez, A., Ramirez-Rojas, A., Chavez-Lopez, R., and Alcocer, J. (2007). Invertebrate assemblages associated with root masses of *Eichhornia crassipes* (Mart.) Solms-Laubach 1883 in the Alvarado Lagoonal System, Veracruz, Mexico. *Aquatic Ecology* **41**, 319–333. doi:10.1007/s10452-006-9054-2

Rost, M. M. (2006). Ultrastructural changes in the midgut epithelium in *Podura aquatica* L. (Insecta, Collembola, Arthropleona) during regeneration. *Arthropod Structure & Development* **35**, 69–76. doi:10.1016/j.asd.2005.10.001

Rougerie, R., Decaens, T., Deharveng, L., Porco, D., James, S. W., Chang, C. H., Richard, B., Potapov, M., Suhardjono, Y., and Hebert, P. D. N. (2009). DNA barcodes for soil animal taxonomy. *Pesquisa Agropecuaria Brasileira* **44**, 789–802. doi:10.1590/S0100-204X2009000800002

Saitou, N., and Nei, M. (1987). The neighbor-joining method – a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.

Salmon, S. (2001). Earthworm excreta (mucus and urine) affect the distribution of springtails in forest soils. *Biology and Fertility of Soils* **34**, 304–310. doi:10.1007/s003740100407

Soto-Adames, F. N. (2002). Molecular phylogeny of the Puerto Rican *Lepidocyrtus* and *Pseudosinella* (Hexapoda: Collembola), a validation of Yoshii's "color pattern species". *Molecular Phylogenetics and Evolution* **25**, 27–42. doi:10.1016/S1055-7903(02)00250-6

Stevens, M. I., and Hogg, I. D. (2006). Contrasting levels of mitochondrial DNA variability between mites (Penthalodidae) and springtails (Hypogastruridae) from the Trans-Antarctic Mountains suggest long-term effects of glaciation and life history on substitution rates, and speciation processes. *Soil Biology & Biochemistry* **38**, 3171–3180. doi:10.1016/j.soilbio.2006.01.009

Stevens, M. I., Greenslade, P., Hogg, I. D., and Sunnucks, P. (2006). Southern Hemisphere springtails: could any have survived glaciation of Antarctica? *Molecular Biology and Evolution* **23**, 874–882. doi:10.1093/molbev/msj073

Stevens, M. I., Porco, D., D'Haese, C. A., and Deharveng, L. (2011). Comment on "Taxonomy and the DNA Barcoding Enterprise" by Ebach (2011). *Zootaxa* **2839**, 85–88.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596–1599. doi:10.1093/molbev/msm092

Torricelli, G., Carapelli, A., Convey, P., Nardi, F., Boore, J. L., and Frati, F. (2010). High divergence across the whole mitochondrial genome in the "pan-Antarctic" springtail *Friesea grisea*: evidence for cryptic species? *Gene (Amsterdam)* **449**, 30–40. doi:10.1016/j.gene.2009.09.006

Tully, T., D'Haese, C. A., Richard, M., and Ferriere, R. (2006). Two major evolutionary lineages revealed by molecular phylogeny in the parthenogenetic collembola species *Folsomia candida. Pedobiologia* **50**, 95–104. doi:10.1016/j.pedobi.2005.11.003