2015-12-15

Development of a barcoding database for the UK Collembola: early results

Shaw, P

Soil Organisms
Senckenberg Museum of Natural History Gorlitz

SOIL ORGANISMS 87 (3) · December 2015 pp. 197–202

Development of a barcoding database for the UK Collembola: early results

Peter Shaw[1]* and Carly M. Benefer[2]

[1] Centre for Research in Ecology, Whitelands College, University of Roehampton SW15 4JD, London, United Kingdom
[2] Plymouth University, School of Biological Sciences, A414 Portland Square, Drake Circus, Plymouth, PL4 8AA, United Kingdom
* Corresponding author, e-mail: p.shaw@roehampton.ac.uk

Received 27 February 2015 | Accepted 9 March 2015
Published online at www.soil-organisms.de 1 December 2015 | Printed version 15 December 2015

Abstract

We report early results from a project to accumulate COI barcodes from UK Collembola to confirm taxonomy and explore their status at an international level. We validated COI sequences for 48 species of Collembola, ranging from 335–670 bp. Of these, seventeen species matched public sequences of the same name, six species were identifiable but the molecular identity disagreed with the morphological identification, and twenty five species gave no reliable match. The successful matches included accurate matches to BINs from countries far from the UK, including Canada, South Africa and Russia. We suggest that, in many cases, these may have been accidentally transported with horticultural materials.

Keywords springtail invasion | taxonomy

1. Introduction

Collembola, or springtails, are ubiquitous in terrestrial systems and numerically the dominant hexapods (typically upwards of 1 animal per cm^2) in many ecosystems (Hopkin 1997). The entire UK species list contains over 400 species, but uncertainty remains both about the many old (probably invalid) names, the continuing arrival of new species and the occurrence of unrecognised cryptic diversity within named species (Shaw et al. 2013). A recent update to the UK Collembola list contained 379 names which have been recorded either reliably or on several occasions, but 169 of these have not been recorded since 2000 (List available at www.Collembola.org). In order to have confidence in any estimates of biodiversity, morphological taxonomy needs to be cross-validated against genetic data. When Porco et al. (2014) used cytochrome *c* oxidase (COI) barcoding to explore the Collembola of a site in Manitoba, Canada, the number of molecular species (97 Molecular Operation Taxonomic Units (MOTUs), using a conservative 14% threshold to delineate species) was more than double the number of morphospecies (45), because of the widespread occurrence of cryptic species. One of the commonest Collembola in Europe, *Parisotoma notabilis* Schäffer, 1896, contains at least 4 clades as genetically distinct from each other as they are from other species in the genus (based on mitochondrial COI and nuclear 28S sequences) (Porco et al. 2012).

Genetic markers have revealed cases where an invasive cryptic species has displaced the native genotype in an otherwise invisible invasion, e.g. the hard rush *Phragmites* (Saltonstall 2002), Tamarisk *Tamarisk* (Gaskin & Schaal 2002) and a tree-feeding adelgid (Havill et al. 2006). We do not yet have a clearly documented case where a non-native genotype of Collembola is demonstrated invading a natural community, but Soto-Adames' (2010) results from Puerto Rico are suggestive of this. Collections on this island in 1927 found the endemic species *Salina walcotti* Folsom (Collembola: Entomobryidae) in several widely spread locations, but it appeared to have become extinct by 1974, displaced by *Salina tristani* Denis, 1933.

ISSN 1864-6417

Soto-Adams showed that Puerto Rico had undergone two distinct invasions by different clades of *Salina tristani* from nearby islands, while the Puerto-Rican endemic *S. walcotti* was re-found in small numbers in one remote isolated mountainous area. Ramirez-Gonzalez et al. (2013) used COI barcoding in a high-throughput system to explore Collembola on the island of Tenerife, and found two pan-European clades (*Parisotoma notabilis* and *Ceratophysella gibbosa* Bagnall, 1940) alongside geographically localised, previously unknown clades of *Friesea truncata* Cassagnau, 1958 that may be a local endemic genotype. A systematic survey of Collembola from the remote island of St Helena found apparently no endemics at all, just common European species, notably *Orchesella cincta* L. which was ubiquitous (Mendel et al. 2008). Genetic work will be needed to confirm the suspicion that these St Helena Collembola are in fact recently introduced mainland clades. Katz et al. (2015) used COI barcodes to explore colour-pattern species within the genus *Entomobrya* in North America and found that 13 colour patterns corresponded to seven genetically isolated lineages. Generally colour was a useful distinguishing feature, but this could only be used post-hoc after confirmation by genetic tests.

There are cases where one species name appears to refer to different Collembola inside and outside the UK. Thus *Folsomia agrelli* Gisin, 1944, defined by its manubrial chaetotaxy, is found in the UK and Scandinavia, but in

Table 1. UK collections of Collembola that match in sequence and name to an existing BIN in the BOLD database.

species	UK Collection site	Genbank accession	BOLD BIN	BOLD collection sites
Allacma fusca (Linneaus, 1758)	Orielton, Wales	KT808323	AAN9178	Canada (3 sites), Estonia, Finland.
Dicyrtomina ornata (Nicolet, 1842)	Bookham, Surrey	KT808331	AAC0663	Normandy, France
Anurida granaria (Nicolet, 1847)	Spadeadam, Northumberland	KT808325	ACS5909	Vestfold, Norway
Folsomia sexoculata (Tullberg, 1871)	Kew, Surrey	KT808382	AAE6252	Canada, Norway
Hypogastrura distincta (Axelson, 1902)	Mold, Wales	KT808350	AAE5828	Russia (Primorsky Krai) – sp. unnamed
Hypogastrura purpurescens (Lubbock, 1867)	Surrey (2 sites)	KT808353	(only on Genbank)	Chile
Isotomurus maculatus (Schäffer, 1896)	Surrey	KT808362	AAC4948	France, South Africa, Marion Island
Isotomurus palustris (Müller, 1776)	Surrey	KT808340	AAO1741	France, Canada (2 sites)
Neanura muscorum (Templeton, 1935)	Northumberland	KT808329	AAT9087	France, Canada
Orchesella cincta (Linnaeus, 1758)	Northumberland	KT808383	AAA6611	Canada (many sites), France, Poland, Moldova
Orchesella villosa (Geoffrey, 1764)	Surrey (2 sites)	KT808342 KT808336	AAA8726	Canada, France, Poland
Parisotoma notabilis Schäffer, 1896	Surrey	KT808351	AAT8983	Canada, France
Pogonognathellus longicornis (Müller, 1776)	Scots highlands + Surrey	KT808377 KT808367	AAW6120	Norwich, UK
Sminthurinus aureus (Lubbock, 1862)	Surrey		AAF3332	Pyrenees, France
Sminthurinus elegans (Fitch, 1863)	Surrey	KT808390	AAB5296 'S. aureus'	France, Canada
Tomocerus minor (Lubbock, 1862)	Surrey	KT808380	AAB5437	France, Canada, Australia
Tomocerus minor (Lubbock, 1862)	Surrey		AAB5440	France, Canada
Xenylla humicola Fabricius, 1780	Lindisfarne	KT808328	AAA6287	Canada

the UK is only found in caves while in mainland Europe is never found in caves, only in mountain soils (Hopkin 2007). The entomobryids *Pseudosinella tarraconensis* Bonet, 1929 and *Pseudosinella dobati* Gisin, 1966, both species of mainland European caves, have been recorded in the UK (Hopkin 2007). However the equation of cave endemics in ancient French caves with similar-looking animals in the UK may not be valid: the surviving specimens are in poor condition and cannot be determined to species, so only a new collection effort can establish the validity of these names in the UK.

Ultimately the only way to establish what Collembola we have in the UK will be to collect widely and aim to define standard sequence data for large numbers of individuals, in conjunction with an international effort to collect equivalent data from as wide a geographical area as is feasible. The sequence that is widely accepted as a de-facto standard for species-level determination of animals is the mitochondrial COI gene (Ratnasingham & Hebert 2007), although problems caused by non-transcribed nuclear copies of this gene (numts; nuclear mitochondrial DNA) (Song et al. 2008) mean that COI-derived conclusions about species boundaries should be validated by co-analyses of a nuclear sequence. Here we report early results from an open-ended programme to collect COI barcodes from UK Collembola, with the dual aims of testing existing UK names against international standards and looking to see where the UK fits in the global distribution of identifiable clades. The intention is to maintain a UK-specific database for Collembola COI barcodes, mirrored on the global BOLD database. We therefore did not focus on one genus in depth, but instead collected widely from the UK fauna on an ad-hoc basis.

2. Methods

Collembola were extracted by Tullgren funnels from leaf litter collected by the first author from multiple locations around the UK, though with a bias towards the south-east. Initially they were collected and stored in 70 % ethanol (based on standard laboratory Industrial Methylated spirit), but this gave poor sequence results and most subsequent sequences came from animals stored in 100 % EtOH in a refrigerator. Species determinations were based on Hopkin (2007), but with confirmation from other sources (Bretfeld 1999, Fjellberg 1998, 2007, Gisin 1960, Jordana 2012, Popatov 2001, Thibeaut et al 2004).

For the molecular identification, DNA was extracted from individuals using a modified ammonium acetate protocol (Benefer 2011) and amplified using a Qiagen Taq PCR Core Kit with modified Folmer primers targeting the COI gene (the standard barcoding region; Ramirez-Gonzalez et al. 2013). PCR products were sent to Macrogen Inc. for Sanger sequencing in the forward direction only. DNA sequences were checked for quality and edited where needed using BioEdit v. 7.0.9.0 (Hall 1999). DNA sequences were also translated to protein using the ExPasy translate tool (http://web.expasy.org/translate/) to ensure their adherence to an open reading frame (Buhay 2009). Where this check failed, chromatograms were re-analysed and sequences re-edited or excluded from further analysis, as appropriate.

DNA sequences were submitted to the BOLD identification system (Ratnasingham & Hebert 2007) to obtain their taxonomic identities, using the Barcode Index Number (BIN) system (Ratnasingham and Hebert, 2013) to assign individuals to Operational Taxonomic Units (OTUs) that closely resemble a species. DNA sequences have been submitted to GenBank (accession numbers listed in Table 1–3).

3. Results

We validated COI sequences for 48 species of Collembola, ranging from 335–670 bp. These sequences varied in length due to varying amplification efficiency and read quality, but the majority of sequences were > 500 bp (43 in total) and only one sequence was discarded following the quality checks. A further thirty-one specimens failed to amplify; most of these had previously been stored at room temperature in 70 % ethanol for a prolonged period. Of the sequences that amplified, 17 species matched sufficiently accurately to public sequences on BOLD to be identifiable as within a BIN (Tab. 1). Six species were identifiable to a BIN, but their identity disagreed with the morphological identification (Tab. 2). Twenty five species gave no reliable match (Tab. 3).

4. Discussion

Although a small dataset, there are some interesting results here. The most noteworthy finding was the global reach of some of the lineages, with the same BIN being detected in the UK and Canada for ten species (Tab. 1). These, plus the records of a UK *Pogonognathellus longicornis* BIN in Victoria (Australia) and a UK *Isotomurus maculatus* BIN from South Africa, possibly result from accidental imports with horticultural/agricultural materials.

Table 2. Species with conflicted identities.

species	UK Collection site	Genbank code	BOLD BIN	BOLD collection sites
Desoria trispinata (Macgillivray, 1896)	Surrey	KT808356	AAA7164	Match to a steel-grey '*Isotoma viridis*' from Pyrenees, France
Pogonognathellus flavescens (Tullberg, 1871)	Scotland	KT808376		99% match to *Pogonognathellus longicornis* but no BIN.
Sminthurinus reticulatus Cassagnau, 1964	Surrey	KT808391		100% match but named '*S. aureus*', '*S. elegans*' and '*Sminthurinus* sp. SA2013'
Sminthurinus domesticus Gisin, 1963	Surrey 2014	KT808384	100% match to *Sminthurinus aureus* and S niger	Surrey 2014
Sminthurinus trinotatus Axelson, 1905	Surrey		Genbank only	97% match to *S. bimaculatus*, France
Tomocerus vulgaris (Tullberg, 1871)	Surrey	KT808364	AAA7970	'*Tomocerus minor*' from France

Table 3. Collections with no matching sequences (as of August 2015).

Species	Collection information	Genbank accession code
Anurida tulbergii Schött, 1891	Surrey 2012	KT808326
Archisotoma pulchella (Moniez, 1890)	Surrey 2012	KT808327
Bilobella braunerae Deharveng, 1978	Shropshire 2015	KT808394
Desoria tigrina Nicolet, 1842	Surrey 2013	KT808337
Dicyrtoma fusca (Lubbock, 1873)	Surrey 2013	KT808355
Folsomia quadrioculata (Tullberg, 1871)	Surrey 2012	KT808344
Friesea claviseta Axelson, 1900	Surrey 2012	KT808354
Heteromurus sp.	Devon (cave) 2014	KT808387
Hypogastrura burkilli (Bagnall, 1940)	Surrey 2012	KT808341
Isotoma viridis Bourlet, 1839	Surrey 2013	KT808360
Isotomurus fucicolus (Schött, 1893)	Surrey 2012	KT808332
Kalaphorura burmeisteri (Lubbock, 1873)	Pengoes, Mid Wales 2012	
Katianna sp. nov.	Cornwall 2012	
Lathriopyga longiseta (Caroli, 1912)	Somerset 2014	KT808385
Monobella grassei (Denis, 1923)	Devon 2013	KT808369
Proisotoma minuta (Tullberg, 1871)	Surrey 2012	KT808348
Protaphorura aurantiaca (Ridley, 1880)	Surrey 2012	KT808371
Pseudisotoma sensibilis (Tullllberg, 1876)	Surrey 2012	KT808346
Pseudosinella alba (Packard, 1873)	Surrey 2012	KT808338
Sminthurides malmgreni (Tullberg, 1876)	Surrey 2012	KT808333
Sminthurides sp.	Kew Gardens Surrey 2014	KT808386
Sminthurinus niger (Lubbock, 1867)	Surrey 2013	KT808361
Tomocerus 'catalanus'	Surrey 2012	KT808352
Tomocerus minor (Lubbock, 1862)	Wales 2012, Scotland 2013	KT808366, KT808379
Xenylla maritima Tullberg, 1869	Surrey 2012	KT808347
Xenylla boerneri Axelson, 1905	Surrey 2012	KT808349

By contrast 25 species provided clear COI sequences that did not correspond to any internationally recognised BINs (Tab. 3), including some common forms for which sequences do exist on BOLD. The implication is that these may be UK-endemic lines, though more collection both in the UK and internationally will be needed to develop confidence in this idea. At least one of these lines is almost certainly an unrecognised import, '*Tomocerus* sp. nov.'. This large springtail resembles *Tomocerus vulgaris* closely (and its closest match on Genbank is a 95% match to *Tomocerus vulgaris*), but the mucro only has 2–3 medial teeth while *T. vulgaris* should have 4–9 such teeth. It also has an unusual red pigmentation of its labrum. Robert Norledge found a springtail matching this description near Reading in 1998, which Steve Hopkin (http://www.stevehopkin.co.uk/collembolamaps/Entomobryomorpha/348.1TOcat/) tentatively called '*T. catalanus*', a Spanish/ southern French tomocerid with 2 medial teeth, but which otherwise does not fit the description.

It is unsurprising that in a minority of cases the apparent morphospecies differed from the name uploaded to BOLD (Tab. 2). In the case of the *Tomocerus vulgaris* that matched a *T. minor,* the only explanation can be a misidentification by one side, as these are clearly separated by the teeth on the dentes. Similarly the two species of *Pogonognathellus* may be distinguished by the filamentous empodium (absent in *P. flavescens*), so these two should not have conflicted. *Desoria trispinata* is a little-known but probably widespread springtail that could easily be mistaken for a grey *Isotoma viridis* and will key there in Hopkin (2007) among other keys (the authors are indebted to Arne Fjellberg, who confirmed the determination of *Desoria trispinata* in the UK).

The other two anomalies come from the genus *Sminthurinus*. *Sminthurinus trinotatus* is visually very similar to *Sminthurinus bimaculatus* (which was a 97% match on Genbank), though they differ in the chaetotaxy of the dens. The distinctively patterned *Sminthurinus reticulatus* (with a ladder-like dorsal pattern) first appeared in the UK in 2006, but has since become common, and a single specimen from Manchester matched 100% to three names on BOLD! These were *Sminthurinus aureus*, *Sminthurinus elegans* and '*Sminthurinus* sp nov.'. The latter two taxa on BOLD included photographs of animals showing a ladder-like dorsal pattern. The simplest explanation here would be that the colour patterns within the '*Sminthurinus aureus*' group are poorly distinguished by COI barcodes so may not represent valid species. More generally, the genus *Sminthurinus* needs a systematic taxonomic effort triangulating genetic, morphometric and colour-pattern data to clarify the species boundaries; the use of a second nuclear marker may aid this process.

There were some issues associated with amplification and quality of the sequence reads produced (which resulted in some sequences < 500 bp), but all of the data submitted was verified according to quality checks similar to those employed by the BOLD BIN analysis (Ratnasingham & Hebert 2013), suggesting the reliability of this data and the assignments to BINs. Although a small and preliminary dataset, it is a valuable starting point for further studies on the taxonomy and community ecology of UK Collembola.

5. Acknowledgements

Thanks are due to the University of Roehampton for the seedcorn funding behind this work and Dr Mairi Knight for facilitating the barcoding work at Plymouth University.

6. References

Benefer, C. M. (2011): The molecular and behavioural ecology of click beetles (Coleoptera: Elateridae) in agricultural land. – University of Plymouth: 274.

Bretfeld, G. (1999): Synopses on Palaearctic Collembola, Vol. 2, Symphypleona. – Abhandlungen und Berichte des Naturkundemuseums Görlitz **71**(1): 1–320.

Buhay, J. E. (2009): "COI-like" sequences are becoming problematic in molecular systematic and DNA barcoding studies. – Journal of Crustacean Biology **29**(1): 96–110.

Fjellberg, A. (1998): The Collembola of Fennoscandinavia and Denmark. Poduridae. – Fauna Entomological Scandinavica **35**, Brill, Leiden.

Fjellberg, A. (2007): The Collembola of Fennoscandinavia and Denmark. Part II: Entomobryomorpha and Symphypleona. – Fauna Entomological Scandinavica **42**, Brill, Leiden.

Gaskin, J. F. & B. A. Schaal (2002): Hybrid Tamarix widespread in US invasion and undetected in native Asian range. – Proceedings of the National Academy of Sciences **99***:* 11256–11259.

Gisin, H. (1960): Collembolenfauna Europas. – Museum d'Histoire Naturelle, Geneva, Switzerland.

Hall, T. (1999): BioEdit: a user-friendly biological sequence alignment editor 220 and analysis program for Windows 95/98/NT. – Nucleic Acids Symposia **41**: 95–98.

Havill, N. P., M. E. Montgomery, G. Yu, S. Shiyake & A. Caccone (2006): Mitochondrial DNA from hemlock woolly adelgid (Hemiptera: Adelgidae) suggests cryptic speciation and pinpoints the source of the introduction to eastern North America. – Annals of the Entomological Society of America **99**: 195–203.

Hopkin, S. P. (1997): Biology of the Springtails. – Oxford University Press, Oxford: 330pp.

Hopkin, S. P. (2007): Key to Collembola (Springtails) of Britain and Ireland. – AIDGAP series (Aids to Identifying Difficult groups of Animals and Plants). – Field Studies Centre, Shrewsbury, UK: 245pp.

Jordana, R. (2012): Synopses on Palaearctic Collembola, Vol. 7/1, Capbryinae and Entomobryini. – Soil Organisms **84**(1): 1–390.

Katz, A. D., G. Giodardno & F. N. Soto-Adames (2015): Operational criteria for cryptic species delimitation when evidence is limited, as exemplified by North American Entomobrya (Collembola: Entomobryidae). – Zoological Journal of the Linnean Society **173**: 818–840 [DOI: 10.1111/zoj.12220].

Mendel, H., P. Ashmole & M. Ashmole (2008*)*: Invertebrates of the central peaks and peak dale, St Helena – Final Report, Saint Helena National Trust.

Popatov, M. (2001): Isotomidae. Synopses on Palaearctic Collembola, Vol. 3. – Abhandlungen und Bereichte des Naturkundemuseums Görlitz **73**(2): 1–603.

Porco, D., D. Skarżyński, T. Decaëns, P. D. N. Hebert & L. Deharveng (2014): Barcoding the Collembola of Churchill: a molecular taxonomic reassessment of species diversity in a sub-Arctic area. – Molecular Ecology Resources **14**: 249–261 [DOI: 10.1111/1755-0998.12172].

Porco, D., M. Potapov, A. Bedos, G. Busmachiu, M. W. Weiner, S. Hamra-Kroua & L. Deharveng (2012): Cryptic Diversity in the Ubiquist Species *Parisotoma notabilis* (Collembola, Isotomidae): A Long-Used Chimeric Species? – PLoS ONE **7**(9): e46056 [DOI: 10.1371/journal.pone.0046056].

Ramirez-Gonzalez R., D. W. Yu, C. Bruce, D. Heavens, M. Caccamo & B. C. Emerson (2013): PyroClean: Denoising Pyrosequences from Protein-Coding Amplicons for the Recovery of Interspecific and Intraspecific Genetic Variation. – PLoS ONE **8**(3): e57615 [DOI: 10.1371/journal.pone.0057615].

Ratnasingham, S. & P. D. N. Hebert (2007): BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). – Molecular Ecology Notes **1**: 355–364.

Ratnasingham, S. & P. D. N. Hebert (2013): A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. – PLoS ONE **8**(7): e66213 [DOI:10.1371/journal.pone.0066213].

Saltonstall, K. (2002): Cryptic invasion by a non-native genotype of the common reed *Phragmites australis* into North America. – Proceedings of the National Academy of Sciences of the USA **99**: 2445–2449.

Shaw, P. J. A, C. Faria & B. C. Emerson (2013): Updating taxonomic biogeography in the light of new methods – examples from Collembola. – Soil Organisms **85**: 161–170.

Song, H., J. E. Buhay, M. F. Whiting & K. A. Crandall (2008): Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are co-amplified. – Proceedings of the National Academy of Sciences **105**: 13486–13491.

Soto-Adames, F. N. (2010): Review of the New World species of *Salina* (Collembola: Paronellidae) with bidentate mucro, including a key to all New World members of *Salina*. – Zootaxa **2333**: 26–40.

Thibeaut, J. M., H. J. Schultze & M. Da Gama Assalino (2004): Synopses on Palaearctic Collembola, Vol. 4, Hypogastruridae. – Abhandlungen und Berichte des Naturkundemuseums Görlitz **75**(2): 1–287.