

Updating taxonomic biogeography in the light of new methods – examples from Collembola

Peter Shaw^{1,*}, Christiana Faria² and Brent Emerson^{2,3}

¹ University of Roehampton, Centre for Research in Ecology, Whitelands College, Holybourne Ave., London SW15 4JD, UK

² School of Biological Sciences, Norwich Research Park, Norwich NR4 7TJ, UK

³ Present Address: Island Ecology and evolution group, IPNA-CSIC, C/Astrofísico Francisco Sánchez 3, 38206 La Laguna, Tenerife, Canary Islands, Spain

* Corresponding author, e-mail: p.shaw@roehampton.ac.uk

Received 10 January 2013 | Accepted 21 October 2013

Published online at www.soil-organisms.de 1 December 2013 | Printed version 15 December 2013

Abstract

The aim of this article is to review the consequences of advances in molecular genetics for researchers of Collembola biogeography. Gene sequence data have generally confirmed existing high-level taxonomy, but complicated species-level taxonomy by uncovering extensive cryptic diversity. Several commonly recorded European Collembola ‘species’ have proved to be complexes of closely related taxa, reducing the value of many older (pre-1990) records to near-zero. It seems likely that many more cryptic species of Collembola remain to be uncovered, even in well studied areas in Europe. The inevitable proliferation of genetically defined ‘species’ will be awkward to integrate into existing databases, which are based on morphospecies. Eventually the results will transform our understanding of Collembola biodiversity, since molecular data contain greatly more information, notably estimates of divergence times. In ancient and relatively pristine ecosystems (e.g. Antarctica, oceanic islands) genetic data can be used to show both extreme isolation of endemics (pre-dating ice ages) and the arrival of European/North American invasives. Collembola from the genus *Lepidocyrtus* in several very different areas show patterns of intra-specific divergence dated in the millions of years, a pattern likely to be repeated in other genera.

An exciting approach to speeding up these genetic studies comes from the advent of high throughput pyrosequencing technology, which already allows mass identification of mixed life forms from clustering then sequencing DNA amplicons (‘metabarcoding’), raising the possibility of automating the identification of bulked soil faunal extracts. A trial of this using soil extracts from Tenerife identified an apparent endemic *Friesea* species, along with mainland invasive species. These next generation techniques rely on Molecular Taxonomic units (MOTUs), identified purely by sequence analysis (usually of mtCOI), so vigilance will always be needed against the spurious identification of sequence artefacts such as nuclear copies of mitochondrial genes (NUMTS). A new risk of relying simply on high throughput sequence analysis is the possibility of inventing non-existent ‘species’ from irrelevant DNA.

Keywords phylobetageography | biodiversity | barcoding | mtCOI | endemism

1. Scope of review

One of the core questions needed for the management and conservation of biodiversity is to know what we have, and where it is (Vane-Wright et al. 1991, Crozier 1992). The recent increase in the availability of genetic

data has been a great asset, for example in the unification of sexually dimorphic species and confirming species boundaries (Janzen et al. 2005, Bik et al. 2012). There is general agreement that genetic data are the pre-eminent tool to resolve the existence of cryptic species in many groups (Hogg & Hebert 2004, Chang et al. 2009, Porco

et al. 2013), but definitions of species boundaries remains contentious. Defining boundaries based on genetic distance for example involves choosing an essentially arbitrary point along a continuum to make a binary division, arguments about its location will inevitably persist. There is a move to a universal standard 'barcode of life' based on the mitochondrial Cytochrome *c* Oxidase 1 (Ratnasingham & Hebert 2007). There are however serious criticisms of the uncritical application of COI barcodes in taxonomy (Ebach & Holdrege 2005, Rubinoff et al. 2006), including the problem of 'NUMTS' (Nuclear MiTochondrial Sequences) (Song et al. 2008). This occurs when mitochondrial DNA sequences are copied into a nuclear chromosome, where they are copied down the generations but never transcribed, hence can accumulate mutations rapidly, unconstrained by selective pressure, and appear as the sequence of an unknown species. In the neotropical butterfly *Astraptes fulgerator* Walch, 1775, initial barcoding by Hebert et al. (2004) suggested that this clade comprises 10 cryptic species (Smith et al. 2006), but Brower (2006) reanalysed the COI data to show that it suggested no more than 7 clades, one of which probably an artefact caused by a NUMTS.

Typically the use of genetic techniques to estimate the extent of evolutionary separation between taxa increases the number of species, or finds deep splits in taxa formerly considered as one species (e.g. Frati et al. 2000, Eggart et al. 2002, Fernando et al. 2003, Blaxter 2003, Timmermans et al. 2005, Janzen et al. 2005, Cicconardi et al. 2009, Cicconardi et al. 2013).

This review will contrast biodiversity of the Collembola patterns based on our current knowledge of morphospecies, and what is starting to be found using DNA sequence data, exemplified using results from three contrasting biomes.

2. Biodiversity within the Collembola

Collembola is an ancient group whose main families evolved before the split-up of Pangea around 200 million years before present (Hopkin 1997). Despite this, Collembola appear not to be very speciose, with recent estimates around 7,900 species globally (Janssens 2013) (up from Deharveng's 2004 figure of c. 7000), compared to more than 300,000 Coleoptera. It is likely that the true species richness of Collembola is much higher: Cicconardi et al. (2013) suggest the true figure for Collembolan species richness is around 500,000, due to the widespread occurrence of cryptic species.

As with other life forms, long-isolated populations of Collembola can give rise to local endemic morphospecies;

e.g. on remote islands (Deharveng 1980, Thibaud 1991, Thibaud & Najt 1993, Fjellberg 1995, Soto-Adames 2002), mountain tops (Dallai 1983, Deharveng & Bedos 1991, Cassagnau 1991) and ancient cave systems (Sendra & Reboleira 2012).

Although collembolan taxonomy at the higher taxonomic levels appears now to be stable (Rusek 2002, Deharveng 2004), at the intrageneric level modern biochemical/genetic techniques have repeatedly confirmed taxonomic splitting and uncovered cryptic species. Thus records hitherto attributed to *Isotoma viridis* (Bourlet, 1839) may now be called *I. viridis*, *I. anglicana* Lubbock, 1862, *I. caerulea* Bourlet, 1839 or *I. riparia* (Nicolet, 1842) (Simonsen et al. 1999, Burkhardt & Filser 2005). The taxon hitherto called *Isotomurus palustris* (Müller, 1776) has been split into 14 species, including (for the UK alone) the original nominate species plus *I. maculatus* (Schäffer, 1896), *I. prasinus* (Reuter, 1891) and *I. unifasciatus* (Börner, 1901) (Carapelli et al. 2001, 2005). The morphological similarity between *Dicyrtomina ornata* (Nicolet, 1842) and *D. saundersi* (Lubbock, 1862) meant that authors from Lubbock (1873) to Hopkin (2007) doubted that they are validly separate species, but molecular data has confirmed the split, at least for mainland Europe (Fanciulli et al. 1995, 2001a). Porco et al. (2013a) attempted to determine Collembolan species richness at Churchill (Manitoba, Canada) by COI barcoding (using a 14% threshold for species separation), and found approximately twice as many genetic species (97) as morphospecies (45). Several common widespread, common Collembola (including *Podura aquatica*, *Folsomia quadrioculata*, *Isotomurus plumosus* and *Sminthurides malmgreni*) proved to conceal multiple Molecular Operational Taxonomic Units, effectively cryptic species.

To be confident that one is dealing with cryptic species one needs to cross-check mitochondrial phylogeny against a nuclear gene sequence, see, e.g., Porco et al. (2013b) who used mtCOI and nuclear 28S to show that the ubiquitous collembolan *Parisotoma notabilis* (Schäffer, 1896) contains 4 genetic lineages each as distinct from each other as any of these '*Parisotoma notabilis*' are from other species in the genus.

The Collembolan genus *Lepidocyrtus* Bourlet, 1839 is a globally distributed genus, widespread in leaf litter (and to a lesser extent on arboreal surfaces) across the northern hemisphere. Current taxonomy of morphospecies within this genus depends heavily on the presence/absence of scales on the antennae and legs, the extent of hyper development of thorax 2 and on colour patterns. Taxonomists have expressed reservations about this, especially the problems of seeing scales on appendages after preservation (Hopkin 2007) and

recent genetic work has confirmed the presence of cryptic diversity within the genus. Cicconardi et al. (2009) collected seven morphospecies of *Lepidocyrtus* around one area of the western Mediterranean and found that two morphospecies [*L. lignorum* (Fabricius, 1775) and *L. curvicollis* Bourlet, 1839] were polyphyletic. '*L. lignorum*' comprised four clades and '*L. curvicollis*' formed two clades, all apparently separating over 10 million years ago. These clades were derived by a mitochondrial sequence (COII) and confirmed using the nuclear Elongation Factor 1 gene (EF1). Subsequently the same genus was studied in Panama, with a similar result: 5 morphospecies of *Lepidocyrtus* using COII and EF1 were found to contain 58 discrete lines with clade separations in excess of 10 million years ago, patterned by the palaeogeography of the region (Cicconardi et al. 2013). Soto-Adames (2002) used COI sequences to check the validity of existing *Lepidocyrtus* taxonomy in Puerto Rico. In most cases the validity of 'colour-pattern species' was upheld, with estimated separation dates in excess of eight million years. It seems an inescapable conclusion that we can expect to find further cryptic species within *Lepidocyrtus*, and that the taxonomy of the genus needs to be revised when sufficient data linking sequence data to morphospecies come available. There is no reason to think that *Lepidocyrtus* is an unusually diverse or ancient one, so it seems likely that many more cryptic species of Collembola remain undetected.

These deep divisions within the genus *Lepidocyrtus* can also be seen as an example of unexpectedly high phylobetadiversity (Emerson et al. 2011), an index of change in genetic composition across spatial gradients, either within species or of communities. The models underlying phylobetadiversity incorporate both history (since ancestral haplotypes may be inferred and their radiations mapped in space) and biotic interactions, making them a powerful new tool for the analysis of difficult communities such as soil fauna (Graham & Fine 2008, Emerson & Gillespie 2008).

Co-occurring with *Lepidocyrtus* spp in many habitats are members of the genus *Entomobrya* Rondani, 1861; both genera are routinely found in arboreal habitats as members of the bark and canopy faunas (Bowden et al. 1976, Shaw et al. 2007). The definitive review of the genus is given by Jordana (2012), though South (1961) is still a useful reference. Taxonomy is based on colour patterns and chaetotaxy, and for collections in the USA existing species boundaries are known to be unreliable, with 'colour morphs' generally being genetically isolated. However there are a minority of cases (e.g. *Entomobrya clitellaria* Guthrie, 1903) where valid species prove to have dimorphic colour patterns (Katz, in prep.).

3. Examples from contrasting biomes

It is instructive to compare the impact of genetic data on our understanding of biodiversity in Collembola from three contrasting biomes: the British Isles, remote oceanic islands and Antarctica. Our knowledge of UK Collembola suffers from the legacy of 150 years of study, since most old records are taxonomically unclear by modern species definitions, generating spurious endemics and unreliable species lists that need DNA sequence data to clarify true affiliations. By contrast remote oceanic islands are little-studied by Collembologists but do harbour endemic species of Collembola, unless displaced by invasive alien soil fauna. In Antarctica the few species of Collembola live in widely isolated populations that pre-date the Pleistocene ice ages, presumably relics of a far richer fauna that was exterminated by massive ice cover.

3.1. The British Isles

From the viewpoint of evolutionary genetics, the British Isles are relatively young, having been repeatedly both connected to the mainland, and covered by ice or periglacial conditions until just ca. 10,000 years bp. (Bowen et al. 2002) They are also one of the best studied countries for many facets of biodiversity and were home to pioneers of Collembola taxonomy, so should have reliable lists and distribution data for the Collembola. In fact this is not the case.

Hopkin (2007) recorded 410 species of Collembola in the UK, but of these 101 names are suspect, defined as being based on a poorly preserved specimen or a probable misidentification. Within the especially problematic genus *Protaphorura* Absolon, 1901, 39 species names have been recorded for the UK (Hopkin 1997), of which only 24 were on the then world list maintained by Bellinger et al. (2009). By 2012 only 15 of these names remained on the world list, of which only 5 were considered to be 'good' species by Pomorski (1998). By contrast, it is probable that when molecular taxonomy is brought to bear on these 5 species of UK *Protaphorura*, they will be found to be hiding cryptic species that correspond poorly or not at all with the previously invalidated names. Hence the biogeography of *Protaphorura* in the UK is currently little known, despite there being many old records. Such bi-directional uncertainty undermines the credibility and accuracy of global species totals (May 1988, Mora et al. 2011). With all these caveats, the current list of accepted Collembola species in the UK is 322, of which 3 are invasives in specialised habitats and probably extinct.

Unlike *Protaphorura*, the genus *Orchesella* Templeton, 1835 is taxonomically uncomplicated and species may

be determined by eye using colour patterns. Distribution data for *Orchesella* species in the UK should therefore be reliable. The commonest and most distinctive is *Orchesella cincta* (Linnaeus, 1758), with 230 UK records of this species dating back to 1834 (Templeton 1835). The UK distribution map (Fig. 1) shows a strong bias towards the London area, with very few records in Cornwall, Wales or northern Scotland. Since the first author of the present review has collected this species with approximately equal ease in all parts of England, Wales and the Scottish lowlands, this distributional bias in fact surely reflects recorder bias (Prendergast et al. 1993), though its apparent absence from high mountains may be accurate, as it is said to be replaced by *Orchesella alticola* Uzel, 1891 (Hopkin 2007). In summary, the main influence visible on the biogeography of this common, large and distinctive species in the UK is recorder bias. For less common or distinctive species this bias can only be worse.

This well-known observer bias problem greatly reduces the usefulness of national records as tools to map distributions, but does not preclude a limited study of temporal change in the distribution of colour-pattern species. Thus the UK has two species of *Dicyrtomina*, *D. ornata* and *D. saundersi*. These can easily be distinguished by the pattern on abd 5 (a solid splotch in *D. ornata* but a multi-lobed pattern in *D. saundersi*). Older sources suggest *D. ornata* is the commoner species, contradicting modern experience, and the distribution of dates of records (Fig. 2) strongly suggests that *D. ornata* has undergone range contraction in the UK while *D. saundersi* has expanded. (Median year of record between the two species are highly significantly different; $p < 0.001$ by a U test). There is no obvious explanation for this, which may be a case of un-noticed competitive displacement, or even the recent displacement of a native line of *D. saundersi* with a more invasive foreign (continental) clade, as seems to have happened with the common reed *Phragmites australis* Trin. ex Streudel in the USA since 1910 (Saltonstall 2002).

The UK apparently supports nine endemic species of Collembola (Table 1), but most of these are *Incertae sedis* (based on inadequate descriptions and ruined type specimens). The only credible of these seems to be *Sminthurinus lawrencei* Gisin, 1963 (with two records and a distinctive mucro). By contrast, *Entomobrya lawrencei* Baquero & Jordana, 2008 (just one collection) is separated from *Entomobrya handschini* Stach, 1922 only by the chaetotaxy of the slide-mounted type specimen (Baquero & Jordana 2008), and the remaining endemics lack a useable type specimen. No UK endemic cave dwelling Collembola have yet been discovered, although ancient cave systems in unglaciated areas may

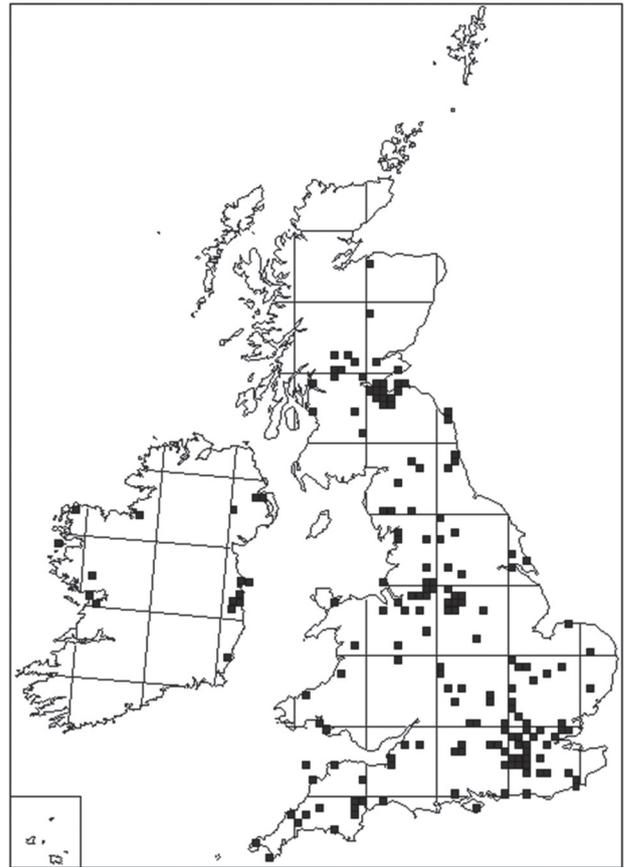


Figure 1. Distribution of records of *Orchesella cincta* in the UK and Ireland, up to 2012.

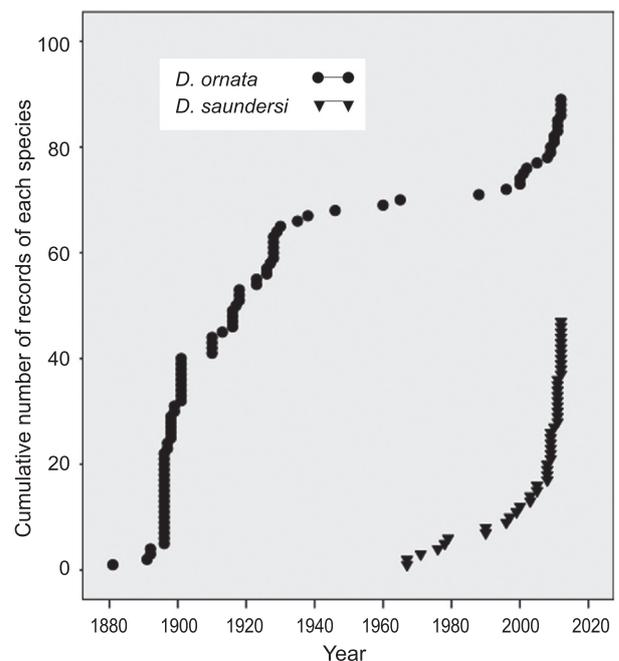


Figure 2. Temporal distribution of records of *Dicyrtomina* species in the UK and Ireland, up to 2012.

Table 1. Collembola apparently endemic to the UK.

Species name	Comment
<i>Anurophorus unguiculus</i> Bagnall, 1940	<i>Incertae sedis</i> . Probably an inadequately described <i>A. satchelli</i> , based on a poorly preserved holotype.
<i>Ballistura ultonica</i> (Carpenter, 1911)	One record from Ireland; <i>Incertae sedis</i> .
<i>Entomobrya lawrencei</i> Baquero & Jordana, 2008	One collection of one animal, by Peter Lawrence in 1964. Previously called <i>E. handschini</i> Stach 1922, but details of the chaetotaxy led Baquero & Jordana (2008) to describe it as a new species.
<i>Folsomia achaeta</i> Bagnall, 1939	<i>Incertae sedis</i> .
<i>Folsomia brevifurca</i> (Bagnall, 1949)	<i>Incertae sedis</i> , type is a juvenile.
<i>Folsomia monoculata</i> (Bagnall 1949)	<i>Incertae sedis</i> , type is a poorly preserved juvenile.
<i>Pachyotoma sphagneticola</i> (Bagnall, 1940)	Only found in one peat bog in Cheshire; Probably juvenile <i>P. crassicauda</i> (Tullberg, 1871).
<i>Protaphorura bagnalli</i> (Salmon, 1959)	One collection, Probably junior synonym of <i>P. armata</i> (Tullberg, 1869).
<i>Sminthurinus lawrencei</i> Gisin, 1963	Described by Gisin from Kent; subsequently collected once in Surrey. Differs from <i>S. niger</i> (Lubbock, 1867) by chaetotaxy of dens and dental pattern on mucro.

contain some (e.g. an undescribed entomobryid has been collected from Lidstone Cave in Devon). In summary, the best explanation for the endemic Collembola in Table 1 is that they are merely taxonomic artefacts, though it is likely that there are UK-endemic Collembola waiting to be discovered and described. The impact of genetic data on our understanding of UK endemism can be guessed as leading to a great increase in the number of UK-endemic clades, though whether these deserve specific status will depend on their distinctiveness.

3.2. Island faunas

Ancient oceanic islands can support high proportions of endemic species, including Collembola. Christiansen & Bellinger (1992) reported 92 Collembola species of the 154 species from Hawaiian Islands to be endemic. They also observed that there was good reason to think that both endemic and established immigrant Collembola had become extinct in the Hawaiian islands in the past 50 years, as had happened with birds and plants (e.g. Loope et al. 1988). There is no proof yet of island-endemic Collembola species having been lost to invasive predators, unlike other invertebrates (Gillespie & Reimer 1993, Ashmole & Ashmole 2000, Coote & Loeve 2003), but there are examples suggesting that losses of endemic Collembola have occurred by ecological displacement. The island of St Helena, being ancient and isolated with many endemic taxa, would have been expected to harbour endemic (or at least exotic) Collembola. However, a systematic collection of Collembola from throughout the island in 2006 found no endemics at all, just a low

diversity set of common European species dominated by 2 large, ubiquitous species *Orchesella cincta* (Linnaeus, 1758) and *Tomocerus minor* (Lubbock, 1862) (Mendel et al. 2008). No genetic data were collected here.

Genetic data about island endemic Collembola faunas is just beginning to appear, suggesting the typical remote island pattern where invaders are brought in accidentally then displace ancient endemics. On Puerto-Rico an entomobryid Collembolan, *Salina wolcottii* Folsom, 1927, was collected from across the island in 1927 but was found to be apparently extinct by 1974 (Soto-Adames 2010). An island-wide *Salina* collecting effort followed by sequencing both nuclear (G6PDH) and mitochondrial (COI) genes identified 3 distinct clusters; the commonest, a species close to *Salina tristani* Denis, 1931, which was widespread throughout the Dominican republic, with a second clade close to *S. tristani* from the Virgin islands, while the native *S. wolcottii* was re-found in one remote mountainous area (Soto-Adames 2010).

The island of Tenerife also showed a mixture of endemics and invasives, with records from two pan-European invasive species of Collembola (*Parisotoma notabilis* (Schäffer, 1896) and *Ceratophysella gibbosa* (Bagnall, 1940) but also two geographically localised surviving endemic lineages (Ramirez-Gonzalez et al. 2013).

3.3. Antarctica

The biogeography of Antarctic Collembola has received a disproportionate amount of study despite (or because) of their isolation, low biodiversity and high percentage

endemism (said to be 70% by Block 1984, but this figure will have changed as new species have been added to the Antarctic list). The consistent finding is that ice sheets are the geographical constraints acting as barriers to dispersal over geological time periods, with populations staying genetically isolated for millions of years. Eastern and western Antarctica contain almost entirely different Collembola, with only one species (*Friesea grisea* Schäffer, 1891) in common (Greenslade 1995). The division between maritime and coastal Antarctica is known as the Gressitt line and is such a sharp biogeographical discontinuity that Chown & Convey (2007) likened it to the Wallace line of south-east Asia. Within localised regions of Antarctica, the same pattern recurs that geographic isolation predicts biogeographical patterns (McGaughan et al. 2010). Thus, a major ice tongue (the Drygalski ice tongue) splits Victoria Land (VL) into two biogeographical zones (Northern VL and Southern VL), with different species of Collembola in the north and south (Sinclair & Stevens 2006). Similarly, the boundary between *Desoria klovstadi* (Carpenter, 1902) and *Gressittacantha terranova* Wise, 1967 is the Mariner glacier (Adams et al. 2006, Stevens et al. 2007). *Gressittacantha terranova* has its population trisected into genetically isolated groups by the Aviator and Campbell glaciers (Fanciulli et al. 2001b, Hawes et al. 2010). In summary the patterns of endemism in Antarctica are dominated by isolation imposed by physical geography.

These findings have been amplified by genetic data, which confirm isolation between remote populations, e.g. Stevens & Hogg (2003) found that *Gomphiocephalus hodgsoni* Carpenter, 1908 (Hypogastruridae) populations on Ross island and Beaufort islands were long-separated from a mainland (Taylor Valley) population (about 2% separation in COI). Antarctic Collembola populations consistently show long periods of divergence, often pre-dating the pleistocene. Mitochondrial sequence data show that the two populations of *Friesea grisea* separated by the Gressitt line show extreme genetic divergence (> 20% in protein coding genes), suggesting prolonged isolation (10–20 million years), pre-dating the development of the Antarctic ice sheet (Torricelli et al. 2010). Similarly, some endemic clades of *Cryptopygus* sp. on the Antarctic mainland were separated by 10–20 million years, suggested to have survived in ice-free refugia since the miocene (Stephens et al. 2006)

There is no evidence yet from either morphological or genetic studies that natural habitats in continental Antarctica have experienced invasions of alien Collembola, presumably due to the inaccessibility and harshness of conditions (Sinclair & Stevens 2006). However colonisation by non-native Collembola has

already started in the maritime Antarctic (Slabber et al. 2007, Greenslade et al. 2012), and at least two non-Antarctic Collembola have been found in McMurdo and Scott Base research stations, in both cases associated with indoor horticultural facilities (Sinclair & Stevens 2006).

4. High throughput techniques

A current standard high-throughput platform, e.g. the Roche 454, allows for highly parallel sample processing. This means that researchers may work with pooled DNA samples (such as from a collection of arthropods whose DNA is co-extracted) and reconstruct the community from first generating clusters of similar amplicons, then identifying them by sequence comparison to known species (Yu et al. 2012). Unlike standard barcode sequences (e.g. the 648 base pairs of cytochrome oxidase 1), 454 amplicons tend to be shorter (100–300 base pairs) for reasons of technical stability. This is a transformative new technology, and consistently finds substantial numbers of unrecognised taxa in environmental collections (e.g. Sogin et al. 2006 for deep sea bacteria, Buée et al. 2009 for fungi, Nolte et al. 2010 for protozoa in lake water, Larsen et al. 2010 for human gut flora, Creer et al. 2010 for meiofauna, Fonseca et al. 2010 for general marine metazoan biodiversity, Porazinska et al. 2010 for rainforest nematodes). The disadvantage of this pooled DNA method is the lack of a collectable specimen of the unrecognised potentially new taxon, whereas DNA sequencing of single specimens can (in principle) be done using a non-destructive extraction method in order to retain the cuticle for morphological analyses (Porco et al. 2010).

Preliminary work applying 454 high-throughput technology to community samples of Collembola on Tenerife successfully reconstructed COI barcodes from the pooled amplicons, confirming the presence of *Parisotoma notabilis* and *Ceratophysella gibbosa* (ubiquitous in synanthropic habitats and found as identical sequences in mainland Europe), along with new sequences close to *Friesea truncata* Cassagnau, 1958 that appeared to be endemic to specific regions of the island (Ramirez-Gonzalez et al. 2013).

Although the field is in its infancy, a logical extension of using high throughput techniques for mapping the genetic diversity of Collembola would be applying the same principle to the amplification and sequencing of environmental DNA (eDNA). Andersen et al. (2012) have largely reconstructed the vertebrate fauna of a safari park by high throughput analyses of eDNA recovered from

the soil. Taberlet et al. (2012) showed that eDNA could be extracted from bulk soil and used to identify both oligochaetes and fungi. qPCR amplification of eDNA in a range of natural ponds streams and lakes in Europe found amplicons from the crustacean *Lepidurus apus* Leach, 1816 and the insect *Leucorrhinia pectoralis* Charpentier, 1825 (Thomsen et al. 2012), along with sequences from fish, mammals and riverside birds. Thus eDNA profiles might prove useful to define aquatic Collembola in the diverse but low-density communities found in freshwater systems such as upwellings (Bketschko & Christian 1989, Shaw et al. 2012) or caves (Deharveng et al. 2008, Sendra & Riboleira 2012). If eDNA extraction from soils and litters can be standardised, it may be possible to replace the expertise-demanding steps of high-gradient extraction, sorting then visual identification of Collembola (and all other soil fauna) by a wholly automated procedure that defines the community by the barcode sequences of its dominant eDNA amplicons. Under this scenario, worries about mis-identification of damaged or juvenile animals will be replaced by worries that unfamiliar sequences are merely NUMTS or amplification artefacts.

5. Conclusions

Our current understanding of Collembolan biogeography, based on morphospecies, is set to be transformed by recent increases in the availability of DNA sequencing. The first effect has been to complicate species distribution maps of the Collembola by revealing an unexpected degree of intra-specific complexity. Generally, though not inevitably, colour pattern ‘species’ have been confirmed to have been genetically isolated, and many literature records for Collembola will become progressively downgraded in value as their morphospecies are found to contain cryptic lineages. The initial consequences of widespread genetic data on our understanding of Collembolan biogeography will be disruptive as old taxa are split, but as data accumulate we will start to see the true biogeography of clades, unhampered by taxonomic uncertainty. In addition to refining our detection of species boundaries, genetic data allow dating of population splits. These have consistently given surprisingly large ages (> 10 MYBP), which correlate with local palaeogeography. Barcoding also allows identification of modern range expansions by aggressive/invasive clades. High throughput techniques raise the possibility of automating the collection of taxonomic barcodes for Collembola (or any other soil arthropod), although this has yet to be demonstrated in practice.

7. Acknowledgements

We are grateful to the Systematics Association and the Linnean Society for SYNTAX funding to undertake barcoding of UK *Entomobrya* and *Lepidocyrtus*. We are also grateful to the anonymous referees who made many helpful comments on the text.

8. References

- Adams, B. J., R. D. Bardgett, E. Ayres, D. H. Wall, J. Aislabie, S. Bamforth, R. Bargagli, C. Cary, P. Cavacini, L. Connell, P. Convey, J. W. Fell, F. Frati, I. M. Hogg, K. K. Newsham, A. O'Donnell, N. Russell, R. D. Seppelt & M. I. Stevens (2006): Diversity and distribution of Victoria Land biota. – *Soil Biology & Biochemistry* **38**: 3003–3018.
- Andersen, K., K. L. Bird, M. Rasmussen, J. Haile, H. Breuning-Madsen, K. H. Kjær, L. Orlando, M. T. P. Gilbert & E. Willerslev (2012): Meta-barcoding of ‘dirt’ DNA from soil reflects vertebrate biodiversity. – *Molecular Ecology* **21**: 1966–1979.
- Ashmole, P. & M. Ashmole (2000): St Helena and Ascension island: a natural history. Anthony Nelson: Oswestry. 528 pp.
- Baquero, E. & R. Jordana (2008): Redescription of *Entomobrya quinquelineata* Börner, 1901 (Collembola: Entomobryidae) and description of three new species. – *Zootaxa* **1821**: 1–12.
- Bellinger, P. F., K. A. Christiansen & F. Janssens (2009): Checklist of the Collembola of the World [http://www.collembola.org] Accessed 29 October 2012.
- Bik H. M., D. L. Porazinska, S. Creer, J. G. Caporaso, R. Knight & W. K. Thomas (2012): Sequencing our way towards understanding global eukaryotic biodiversity. – *Trends in Ecology and Evolution* **27**: 233–243.
- Bketschko, G. & E. Christian (1989): Collembola in the bed sediments of an Alpine gravel stream (RITRODAT-Lunz study area, Austria): – *Internationale Revue der gesamten Hydrobiologie und Hydrographie* **74**: 491–498.
- Blaxter, M. (2003): Molecular systematics: Counting angels with DNA. – *Nature* **421**: 122–124.
- Block, W. (1984): Terrestrial Ecology: invertebrates and ecosystems. – *Antarctic Ecology* **1**: 163–236.
- Bowden, J., I. H. Haines & D. Mercer (1976): Climbing Collembola. – *Pedobiologia* **16**: 298–312.
- Bowen, D. Q., F. M. Phillips, A. M. McCabe, P. C. Knutz & G. A. Sykes (2002): New data for the Last Glacial Maximum in Great Britain and Ireland. – *Quaternary Science Reviews* **21**: 89–101.
- Brower, A. V. Z. (2006): Problems with DNA barcodes for species delimitation: ‘ten species’ of *Astraptus fulgerator* reassessed (Lepidoptera: Hesperidae): – *Systematics and Biodiversity* **4**: 127–132.

- Buée, M., M. Reich, C. Murat, E. Morin, R. H. Nilsson, S. Uroz & F. Martin (2009): 454 pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. – *New Phytologist* **184**: 449–456.
- Burkhardt U. & J. Filser (2005): Molecular evidence for a fourth species within the *Isotoma viridis* group (Insecta, Collembola). – *Zoologica Scripta* **34**: 177–185.
- Carapelli, A., F. Frati, P. Fanciulli & R. Dallai (2001): Taxonomic revision of 14 south-western European species of *Isotomurus* (Collembola, Isotomidae), with description of four new species and the designation of the neotype for *I. palustris*. – *Zoologica Scripta* **30**: 115–143.
- Carapelli, A., F. Frati, P. Fanciulli, F. Nardi & R. Dallai (2005): Assessing species boundaries and evolutionary relationships in a group of south-western European species of *Isotomurus* (Collembola: Isotomidae) using allozyme data. – *Zoologica Scripta* **34**: 71–79.
- Cassagnau, P. (1991): *Camerounura* n.g. n.sp. Collembola Neanurinae endémique du Mt Cameroun. – *Revue d'Ecologie et Biologie du Sol* **28**: 221–224.
- Chang, C. H., R. Rougerie & J. H. Chen (2009): Identifying earthworms through DNA barcodes: pitfalls and promise. – *Pedobiologia* **52**: 171–180.
- Chown S. L. & P. Convey (2007): Spatial and temporal variability across life's hierarchies in the terrestrial Antarctic. – *Philosophical Transactions of the Royal Society B* **362**: 2307–2331.
- Christiansen, K. & P. Bellinger (1992): *Insects of Hawaii: Volume 15, Collembola*. – University of Hawaii Press, Honolulu: 456 pp.
- Cicconardi, F., P. P. Fanciulli & B. C. Emerson (2013): Collembola, the biological species concept, and the underestimation of global species richness. – *Molecular Ecology* [doi:10.1111/mec.12472].
- Cicconardi, F., F. Nardi, B. C. Emerson, F. Frati & P. P. Fanciulli (2009): Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. – *Molecular Ecology* **19**: 386–400.
- Coote, E. & E. Loeve (2003): From 61 to 5; endemic tree snails of the Society Islands fall prey to an ill judged biological control programme. – *Oryx* **37**: 91–96.
- Creer, S., V. G. Fonseca, D. L. Porazinska, R. M. Giblin-Davis, W. D. Sung, D. Power, M. Packer, G. R. Carvalho, M. L. Blaxter, P. J. D. Lamshead & W. K. Thomas (2010): Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. – *Molecular Ecology* **19** (suppl. 1): 4–20.
- Crozier, R. H. (1992): Genetic diversity and the agony of choice. – *Biological Conservation* **61**: 11–15.
- Dallai, R. (1983): Interesse biogeografico dei Neanuridi (Collembola) della Sardegna e delle isole dell'Archipelago Toscano. – *Lavori della Società Italiana di Biogeografia* **8**: 417–465.
- Deharveng, L. (1980): Contribution a la connaissance systemique, ecologique et biogeographique des Collemboles Neanurinae de Corse. – *Revue d'écologie et de biologie du sol* **17**: 419–435.
- Deharveng, L. (2004): Recent advances in Collembola systematics. – *Pedobiologia* **48**: 415–433.
- Deharveng, L. & A. Bedos (1991): Taxonomy and cladistics analyses of the Thai species of *Friesea* Dalla Torre (Collembola: Neanuridae). – *Tropical Zoology* **4**: 287–315.
- Deharveng, L., C. D'Haese & A. Bedos (2008): Global diversity of springtails (Collembola: Hexapoda) in freshwater. – *Hydrobiologia* **595**: 329–338.
- Ebach, M. C. & C. Holdrege (2005): DNA barcoding is no substitute for taxonomy. – *Nature* **434**: 697.
- Eggart, L. S., C. A. Rasner & D. S. Woodruff (2002): The evolution and phylogeography of the African elephant inferred from mitochondrial DNA sequence and nuclear microsatellite markers. – *Proceedings of the Royal Society B* **269**: 1993–2006.
- Emerson, B. C. & R. Gillespie (2008): Phylogenetic analysis of community assembly and structure over space and time. – *Trends in Ecology and Evolution* **23**: 619–630.
- Emerson, B. C., F. Cicconardi, P. P. Fanciulli & P. J. A. Shaw (2011): Phylogeny, phylogeography, phylobetadiversity and the molecular analysis of biological communities. – *Philosophical Transactions of the Royal Society Series B* **366**: 2391–2402.
- Fanciulli, P. P., F. Frati, A. Carapelli & R. Dallai (1995): Genetic diversity within and between populations of *Dicyrtomina ornata* and *Dicyrtomina saundersi* (Collembola, Dicyrtomidae). – *Polskie Pismo Entomologiczne* **64**: 21–28.
- Fanciulli, P. P., D. Summa, R. Dallai & F. Frati (2001b): High levels of genetic variability and population differentiation in *Gressittacantha terranova* (Collembola, Hexapoda) from Victoria Land, Antarctica. – *Antarctic Science* **13**: 246–254.
- Fanciulli, P. P., M. Belloni, A. Carapelli, R. Dallai & F. Frati (2001a): Taxonomic diagnosis of *Dicyrtomina ornata* and *D. saundersi* (Collembola: Dicyrtomidae) and analysis of their population genetic structure. – *Zootaxa* **10**: 1–16.
- Fernando, P., T. N. C. Vidya, J. Payne, M. Stuewe, G. Davison, R. J. Alfred, P. Andae, E. Bosi, A. Kilbourn & D. J. Melnick (2003): DNA Analysis Indicates That Asian Elephants Are Native to Borneo and Are Therefore a High Priority for Conservation. – *Plos Biology* **1** [doi:10.1371/journal.pbio.0000006].
- Fjellberg, A. (1995): The systematic position of the monotypic family Isotogastruridae (Collembola) with description of *Isogastrura coronata* n. sp. from Fuerteventura, Canary Islands. – *Miscellanea Zoologica* **17**: 123–127.
- Fonseca, V. G., G. R. Carvalho, W. Sung, H. F. Johnson, D. M. Power, S. P. Neill, M. Packer, M. L. Blaxter, P. J. D. Lamshead, W. K. Thomas & S. Creer (2010): Second-generation environmental sequencing unmasks marine metazoan biodiversity. – *Nature Communications* **1**: 1–8.
- Frati, F., E. Dell'Amplio, S. Casasanta, A. Carapelli & P. P. Fanciulli (2000): Large Amounts of Genetic Divergence among Italian Species of the Genus *Orchesella* (Insecta, Collembola) and the

- Relationships of Two New Species. – *Molecular Phylogenetics and Evolution* **17**: 456–461.
- Gillespie, R. G. & N. Reimer (1993): The effects of alien predatory ants on Hawaiian endemic spiders (Aranea: Tetragnathidae). – *Pacific Science* **47**: 21–33.
- Graham, C. H. & P. V. A. Fine (2008): Phylogenetic beta diversity: linking ecological and evolutionary processes over space and time. – *Ecological Letters* **11**: 1265–1277.
- Greenslade, P. (1995): Collembola from the Scotia Arc and Antarctic Peninsula including descriptions of two new species and notes on biogeography. – *Polskie Pismo Entomologiczne* **64**: 305–319.
- Greenslade, P., M. Popatov, D. Russell & P. Convey (2012): Global Collembola on deception island. – *Journal of Insect Science* **12**: 1–16.
- Hawes, T. C., G. Torricelli & M. I. Stevens (2010): Haplotype diversity in the Antarctic springtail *Gressittacantha terranova* at fine spatial scales – a Holocene twist to a Pliocene tale. – *Antarctic Science* **22**: 766–773.
- Hebert, P. D. N., E. H. Penton, J. M. Burns, D. H. Janzen & W. Hallwachs (2004): Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. – *Proceedings of the National Academy of Sciences* **101**: 14812–14817.
- Hogg, I. D. & P. D. N. Hebert (2004): Biological identification of springtails (Hexapoda: Collembola) from the Canadian Arctic, using mitochondrial DNA barcodes. – *Canadian Journal of Zoology* **82**: 749–754.
- Hopkin, S. P. (1997): *Biology of the Springtails*. – OUP, Oxford: 330 pp.
- Hopkin, S. P. (2007): *Key to Collembola (Springtails) of Britain and Ireland*. – AIDGAP, Shrewsbury, UK: 245 pp.
- Janssens, F. (2013): www.Collembola.org, accessed July 2013.
- Janzen, D. H., M. Hajibabaei, J. M. Burns, W. Hallwachs, E. Remigio & P. D. N. Hebert (2005): Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. – *Philosophical Transactions of the Royal Society Series B* **360**: 1835–1845.
- Jordana, R. (2012): Synopses on Palaearctic Collembola 7/1 - Capbryinae & Entomobryini. – *Soil Organisms* **84** (1): 1–390.
- Larsen, N., F. K. Vogensen, F. W. van den Berg, D. S. Nielsen, A. S. Andreasen, B. K. Pedersen, W. A. Al-Soud, S. J. Sørensen, L. H. Hansen & M. Jakobsen (2010): Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. – *PLoS One* **5**: e9085.
- Loope, L. L., O. Hamman & C. P. Stone (1988): Comparative conservation biology of oceanic archipelagoes. – *Bioscience* **38**: 272–282.
- Lubbock, J. (1873): *Monograph of the Collembola and Thysanura*. – Ray Society, London: 466 pp.
- May, R. M. (1988): How many species are there on earth? – *Science* **247**: 1441–1449.
- McGaughan, A., M. I. Stevens & B. R. Holland (2010): Biogeography of circum-Antarctic springtails. – *Molecular Phylogenetics and Evolution* **57**: 48–58.
- McGaughan, A., G. Torricelli, A. Carapelli, F. Frati, M. I. Stevens, P. Convey & I. D. Hogg (2010): Contrasting phylogeographical patterns for springtails reflect different evolutionary histories between the Antarctic Peninsula and continental Antarctica. – *Journal of Biogeography* **37**: 103–119.
- Mendel, H., P. Ashmole & M. Ashmole (2008): *Invertebrates of the central peaks and peak dale, St Helena – Final Report*. Saint Helena National Trust.
- Mora, C., D. P. Tittensor, S. Adl, A. G. B. Simpson & B. Worm (2011): How many species are there on earth and in the ocean? – *PLoS Biol* **9**(8): e1001127.
- Nolte, V., R. V. Pandey, S. Jost, R. Medinger, B. Ottenwalder, J. Boenigk & C. Schlotterer (2010): Contrasting seasonal abundance between rare and abundant taxa conceals the extent of protist diversity. – *Molecular Ecology* **19**: 2908–2915.
- Pomorski, R. J. (1998): *Onychiurinae of Poland (Collembola: Onychiuridae)*. – Wrocław, Poland: Polish taxonomical society: 201 pp.
- Porazinska, D. L., R. M. Giblin-Davis, A. Esquivel, T. O. Powers, W. Sung, W. K. Thomas (2010): Ecometagenetics confirm high tropical rainforest nematode diversity. – *Molecular Ecology* **19**: 5521–5530.
- Porco, D., R. Rougerie, L. Deharveng & P. D. N. Hebert (2010): 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**: 942–5.
- Porco, D., D. Skarżyński, T. Decaëns, P. D. N. Hebert & L. Deharveng (2013a): Barcoding the Collembola of Churchill: a molecular taxonomic reassessment of species diversity in a sub-Arctic area. – *Molecular Ecology Resources* [doi: 10.1111/1755-0998.12172].
- Porco, D., M. Potapov, A. Bedos, G. Busmachiu, M. W. Weiner, S. Hamra-Kroua & L. Deharveng (2013b): Cryptic Diversity in the Ubiquitous Species *Parisotoma notabilis* (Collembola, Isotomidae): A Long-Used Chimeric Species? – *PLoS One* **7**: e46056.
- Prendergast, J. R., R. M. Quinn, J. H. Lawton, B. C. Eversham & D. W. Gibbons (1993): Correcting for variation in recording effort in analyses of diversity hotspots. – *Biodiversity Letters* **1**: 39–53.
- 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**: e57615.
- Ratnasingham, S. & P. D. N. Hebert (2007): BOLD: The Barcode of Life Data System [<http://www.barcodinglife.org>]. – *Molecular Ecology Notes* **1**: 355–364.

- Rubinoff, D., S. Cameron & W. Kipling (2006): A Genomic Perspective on the Shortcomings of Mitochondrial DNA for 'Barcoding' Identification. – *Journal of Heredity* **97**: 581–594.
- Rusek, J. (2002): Taxonomy of Collembola at the beginning of a new millennium. – *Pedobiologia* **46**: 215–224.
- 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.
- Sendra, A. & A. S. Reboleira (2012): The world's deepest subterranean community - Krubera-Voronja Cave (Western Caucasus). – *International Journal of Speleology* **41**: 221–230.
- Shaw P. J. A., M. Dunscombe & A. Robertson (2012): Collembola in the hyporheos of a karstic river: a new Collembolan community containing a new genus for the UK. – *Soil Organisms* **83** (3): 507–514.
- Shaw, P. J. A., C. M. P. Ozanne, M. Speight & I. Palmer (2007): Edge effects and arboreal Collembola in coniferous plantations. – *Pedobiologia* **51**: 287–293.
- Simonsen, V., J. Filser, P. H. Krogh & A. Fjellberg (1999): Three species of *Isotoma* (Collembola, Isotomidae) based on morphology, isozymes and ecology. – *Zoologica Scripta* **28**: 281–287.
- Sinclair, B. J. & M. I. Stevens (2006): Terrestrial microarthropods of Victoria Land and Queen Maud Mountains, Antarctica: Implications of climate change. – *Soil Biology and Biochemistry* **38**: 3158–3170.
- Slabber, S., M. R. Worland, H. P. Leinaas & S. L. Chown (2007): Acclimation effects on thermal tolerances of springtails from sub-Antarctic Marion Island: Indigenous and invasive species. *Journal of Insect Physiology* **53**: 113–125.
- Smith, M. A., N. E. Woodley, D. H. Janzen, W. Hallwachs & P. D. N. Hebert (2006): DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). – *Proceedings of the National Academy of Sciences* **103**: 3657–3662.
- Sogin, M. L., H. G. Morrison, J. A. Huber, D. M. Welch, S. M. Huse, P. M. Neal, J. M. Arrieta & G. J. Herndl (2006): Microbial diversity in the deep sea and the underexplored 'rare biosphere'. – *Proceedings of the National Academy of Sciences* **103**: 12115–12120.
- 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. (2002): Molecular phylogeny of the Puerto Rican *Lepidocyrtus* and *Pseudosinella* (Hexapoda: Collembola), a validation of Yoshii's "colour pattern species". – *Molecular Phylogenetics and Evolution* **25**: 27–42.
- 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.
- South, A. (1961): The taxonomy of the British species of *Entomobrya*. – *Transactions of the Royal Entomological Society* **113**: 87–416.
- Stevens, M. I. & I. D. Hogg (2003): Long-term isolation and recent range expansion revealed for the endemic springtail *Gomphiocephalus hodgsoni* from southern Victoria Land, Antarctica. – *Molecular Ecology* **12**: 2357–2369.
- Stevens M. I., P. Greenslade, I. D. Hogg & P. Sunnucks (2006): Southern Hemisphere Springtails: Could Any Have Survived Glaciation of Antarctica? – *Molecular Biology and Evolution* **23**: 874–882.
- Stevens, M. I., F. Frati, A. McGaughran, G. Spinsanti & I. D. Hogg (2007): Phylogeographic structure suggests multiple glacial refugia in northern Victoria Land for the endemic Antarctic springtail *Desoria klovstadi* (Collembola, Isotomidae). – *Zoologica Scripta* **36**: 201–212.
- Taberlet, P., S. M. Prud'Homme, E. Campione, J. Roy, C. Miquel, W. Shehzad, L. Gielly, D. Rioux, P. Choler, J.-C. Clément, C. Melodelima, F. Pompanon & E. Coissac (2012): Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. – *Molecular Ecology* **21**: 1816–1820.
- Templeton, R. (1835): Descriptions of the Irish species of Thysanura. – *Transactions of the Entomological Society* **1**: 89–98.
- Thibaud, J. M. (1991): Les Collemboles des Petits Antilles. V. Considerations biogéographiques et écophysiologicals sur les Collemboles Poduromorphs. – *Revue d'écologie et de biologie du sol* **28**: 91–111.
- Thibaud, J. M. & J. Najt (1993): Les Collemboles (Insecta) de l'atoll de Fangataufa. – *Bulletin du Museum National d'Histoire naturelle Paris, Series 4*, **15A**: 95–105.
- Thomsen, P. F., J. Kielgast, L. L. Iversen, C. Wiuf, M. Rasmussen, M. T. P. Gilbert, L. Orlando & E. Willerslev (2012): Monitoring endangered freshwater biodiversity using environmental DNA. – *Molecular Ecology* **21**: 2565–2573.
- Timmermans, M. J. T. N., J. Ellers, J. Mariën, S. C. Verhoef, E. B. Ferwerda & N. M. van Straalen (2005): Genetic structure in *Orchesella cincta* (Collembola): strong subdivision of European populations inferred from mtDNA and AFLP markers. *Molecular Ecology* **14**: 2017–2024.
- Torricelli, G., A. Carapelli, P. Convey, F. Nardi, J. L. Boore & F. Frati (2010): High divergence across the whole mitochondrial genome in the 'pan-Antarctic' springtail *Friesea grisea*: Evidence for cryptic species? – *Gene* **449**: 30–40.
- Vane-Wright, R. I., C. J. Humphries & P. H. Williams (1991): What to protect?—Systematics and the agony of choice. – *Biological Conservation* **55**: 235–254.
- Yu, D. W., Y. Ji, B. C. Emerson, X. Wang, C. Ye, C. Yang & Z. Ding (2012): Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. – *Methods in Ecology and Evolution* **3**: 613–623.