Soil Biology and Biochemistry 125 (2018) 37–43

Contents lists available at ScienceDirect

Soil Biology and Biochemistry

journal homepage: www.elsevier.com/locate/soilbio

A DNA metabarcoding approach to characterize soil arthropod communities

Angela M. Oliverio[a,b], Huijie Gan[c], Kyle Wickings[c], Noah Fierer[a,b,*]

[a] Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, 80309, United States
[b] Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, 80309, United States
[c] Department of Entomology, New York State Agricultural Experiment Station, Cornell University, Geneva, NY, 14456, United States

ARTICLE INFO

Keywords:
Soil DNA
Soil arthropods
Morphological identification
Environmental DNA sequencing
CO1 metabarcoding

ABSTRACT

Belowground arthropod communities are diverse and our ability to characterize them remains logistically difficult and time consuming. Molecular metabarcoding techniques are routinely used to assess the diversity of both microbial and some 'macrobial' taxa across a range of environments, but the use of such techniques for characterizing soil arthropod diversity remains limited. Here we used three approaches to profile soil arthropod communities at the family level of resolution across 10 distinct sites via morphological identification, metabarcoding of DNA from the extracted arthropods, and metabarcoding directly from bulk soils. Although the three methods differed to some degree in their ability to detect some individual taxa, we found that all three methods yielded well-correlated site-level estimates of diversity (Spearman's $\rho \geq 0.63$ with $P < 0.05$ for all correlations) and overall arthropod community composition (Mantel $\rho \geq 0.45$ with $P < 0.05$). Of particular note is that DNA extracted directly from bulk soil yielded results comparable to analyses of DNA from extracted arthropods. Thus, DNA metabarcoding of bulk soil will likely be a useful tool for those researchers looking to incorporate multi-domain comparisons or for studies that require rapid assessments of arthropod diversity across a large number of soil samples.

1. Introduction

Soil fauna, including nematodes, annelids, and arthropods (e.g. mites, springtails, centipedes) are key contributors to the functioning of ecosystems with the biomass of soil animals typically representing 40–80% of the total animal biomass found in ecosystems (both aboveground and belowground combined; Fierer et al., 2009). The arthropods are particularly ubiquitous and diverse in soil: a 1 m square plot may host hundreds to thousands of soil arthropod species (Schaefer and Schauermann, 1990). However much of the faunal diversity in soil remains poorly described, making the quantification of faunal diversity and identification of specific taxa difficult (Wall et al., 2005).

Progress in soil arthropod research has been constrained by methodological challenges (André et al., 2002). The morphological identification of soil arthropods is time consuming (Bienert et al., 2012; Querner and Bruckner, 2010) and usually requires significant taxonomic expertise as the diversity is often difficult to characterize due to morphological ambiguity and cryptic diversity (e.g. André et al., 2001; Smith et al., 2008; Yu et al., 2012). Further, there are known biases in the methods commonly used for extraction of arthropod communities (André et al., 2002; Edwards, 1991) and extraction efficiency varies for different taxa across soil types and under different extraction conditions (André et al., 2002; Macfadyen, 1962). These challenges can constrain efforts to characterize soil arthropod communities, particularly when seeking to analyze a larger number of soils in a relatively short period of time or when taxonomic expertise is lacking (Querner and Bruckner, 2010). For example, rapid surveys or environmental assessments, such as the biomonitoring of pests or invasive species; or large-scale studies investigating how soil arthropod communities are distributed in time and space and how they respond to global change factors are often difficult to execute due to these methodological challenges.

While metabarcoding techniques have been widely applied to investigate soil microbial communities (e.g. Bates et al., 2013; Delgado-Baquerizo et al., 2018; Tedersoo et al., 2014) the use of these techniques to survey soil arthropod communities remains limited. While previous studies have successfully used metabarcoding techniques to assess the diversity of specific lineages such as springtails (Hogg and Hebert, 2004), nematodes (Griffiths et al., 2006; Read et al., 2006; Waite et al., 2003), and earthworms (Bienert et al., 2012; Porco et al., 2013), relatively few studies have tested the fidelity of metabarcoding of soil arthropods for community-scale analyses. Soil nematode communities were recently profiled across a range of soil types (Griffiths et al., 2018; Treonis et al., 2018), although both studies noted significant discrepancies between the molecular and morphology-based

* Corresponding author. Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, 80309, United States.
E-mail address: noah.fierer@colorado.edu (N. Fierer).

https://doi.org/10.1016/j.soilbio.2018.06.026
Received 22 March 2018; Received in revised form 25 June 2018; Accepted 27 June 2018

assessments which require further standardization. Although a handful of studies have sequenced regions of the 18S rRNA gene from extracted soil DNA to characterize soil fauna (Wu et al., 2011), the evaluations of these molecular-based methods have found significant differences between the molecular and morphological profiles (Hamilton et al., 2009; Wu et al., 2009). However, DNA metabarcoding has been well-validated for characterizing aboveground arthropod communities (Madden et al., 2016), including arthropods that accumulate in pitfall traps (Ji et al., 2013; Yang et al., 2014).

In this study, we used three approaches to characterize and compare soil arthropod communities at the order and family level: morphological identification from Berlese funnel extraction, metabarcoding of DNA from extracted arthropods, and metabarcoding of DNA from bulk soil. We chose to identify invertebrates to the order and family levels because these taxonomic levels represent a compromise between taxonomic detail and known ecological function and are commonly used for bioindicator and other soil ecological studies (Gan and Wickings, 2017; Gergócs and Hufnagel, 2009; Gulvik, 2007; Ruf, 1998; Socarrás, 2013; Wickings and Grandy, 2013). Further, many soil arthropods can only be classified with confidence to the family level of resolution using readily available taxonomic keys and minimal slide preparation for morphological identification.

We tested the efficacy of the three approaches across a range of soils from ten sites that varied with respect to vegetation type and management intensity. Our objectives were twofold. First, we wanted to compare a metabarcoding approach (high-throughput sequencing of a portion of the mitochondrial CO1 gene, a 'barcode' commonly used for characterization of arthropod diversity, Madden et al., 2016) versus a more traditional, morphology-based approach for quantifying the structure of soil arthropod communities. Second, we wanted to determine whether the metabarcoding approach can be used with bulk soil DNA, bypassing the need to first extract arthropods from soil using standard approaches (e.g. Berlese funnels) as this method is still time intensive and efforts to extract arthropods from soil can introduce significant biases. We addressed these objectives by comparing how the three approaches differed with respect to their estimates of standard ecological metrics: detection of specific arthropod families, community richness, and overall arthropod community composition.

2. Materials and methods

2.1. Site description and soil sampling

We collected surface (0–5 cm) soil samples from ten sites near Ithaca, New York. The ten sites sampled were diverse, spanning a range of vegetation types (grass, forest, crops) and soil management intensities (from sites with no management, including designated natural areas and state forests, to heavily managed sites, including apple orchards and golf courses). For detailed site information, see Supplementary Table 1.

At each site, we randomly selected four 1 m by 1 m plots at least 100 m apart. Four pairs of soil cores of 5 cm × 5 cm were collected from each plot after removing plant litter from the soil surface. Four soil cores from each plot (one core from each pair) were bulked, resulting in four composited soil samples per site for extraction of soil arthropods using Berlese funnels and subsequent morphological and DNA-based identification of the extracted arthropods. The remaining four soil cores from the same plot were combined for bulk soil DNA extraction. As such, a total of eight composited soil samples were collected from each site for comparative identification by morphology and CO1 metabarcoding: yielding a total of 80 soil samples collected for the study. All soil samples were transferred to the lab in coolers, with the soil samples collected for DNA extraction immediately frozen and stored at −20 °C prior to extraction. Importantly, we note that our goal was not to exhaustively census all arthropod taxa found in each of the 10 sites, rather we used these four composited samples per site to broadly survey the dominant arthropod communities from diverse habitats and evaluate different methodologies for characterizing these communities.

2.2. Soil extractions and identification by morphology

We compared arthropod community composition and diversity across the ten sites using three methods: morphological identification of funnel-extracted arthropods, CO1 gene sequencing of funnel-extracted arthropods, and CO1 sequencing of DNA extracted directly from bulk soil samples. We refer to these three methods in brief as 'morph ID', 'arthropod DNA', and 'bulk soil DNA', respectively. To extract arthropods for morphological identification, bulked soil from each plot was placed on Berlese funnels for heat extraction, with the temperature directly above the soil sample starting at 30 °C and increased in 5 °C increments daily reaching a final temperature of 50 °C. Organisms that passed through the funnels were collected in 95% ethanol in plastic cups which were placed directly underneath the funnels. Arthropods were counted under a dissecting microscope at 50× magnification and identified to family level. Slide mounting and examination under a compound microscope at 200x magnification was necessary for accurate identification of some organisms. The following keys were used: Krantz and Walter (2009) for mites; Christiansen and Janssens (2010) for springtails; and Triplehorn and Johnson (2005) for insects; Zhang (1998) for myriadpods; Buddle (2010) for pseudoscorpions; and Kaston (1978) for Araneae. Around 1–2% of the individuals collected were not identifiable due to damage in the key structure of the specimens or small body size (< 0.1 mm). After sorting, the specimens were stored in 95% ethanol before DNA extraction, which was conducted within 4 months of sample collection.

2.3. Identification by CO1 barcodes

To characterize the soil arthropod communities via barcode sequencing, we first extracted DNA from bulk soils. We homogenized 10 g of soil with liquid N_2 and extracted DNA from 0.2 g duplicate subsamples from each of the 40 samples as per Oliverio et al. (2017), with an additional subsample extracted from 10 of the soil samples (one from each site), yielding 90 soil DNA extractions in total. We also pooled all of the preserved arthropods (including any soil falling through the Berlese funnel during extraction) that were used for the morphological identifications by plot (n = 40) and then extracted DNA using the PowerSoil DNA extraction kit (MoBio Laboratories Inc.). We ran multiple extraction blanks to check for potential contamination.

We then amplified a portion of the CO1 gene with arthropod-specific primers as per Madden et al., 2016. Briefly, we PCR amplified a ~158 bp region of the mitochondrial cytochrome *c* oxidase subunit I gene with arthropod-specific primers (Zeale et al., 2011). The primers were modified to permit multiplex sequencing on the Illumina MiSeq platform with the appropriate Illumina adapters included on both primers with the reverse primers also having an error-correcting 12-base pair barcode unique to each sample to permit multiplexed sequencing (Hamady et al., 2008). Duplicate PCR reactions were conducted on all extracted DNA samples along with multiple 'no template' negative controls per plate to check for contamination. Amplicon concentrations were standardized using SequalPrep Normalization plate kits (Invitrogen) and then sequenced on the Illumina MiSeq platform at the University of Colorado Next Generation Sequencing Facility with the 2 × 150 base pair (bp) paired-end chemistry.

We generated approximately 8 million sequence reads across 170 samples including 80 from extracted invertebrates and 90 from bulk soil. Sequences were demultiplexed with the custom Python script 'prep_fastq_for_uparse_paired.py' (Leff, 2016) and then sequences were merged, quality filtered, and clustered into phylotypes with UPARSE (Edgar, 2013). We merged forward and reverse reads, retaining 3.7 million reads (47%). Next, reads were discarded if the paired end reads did not have a minimum overlap of at least 16 bp or if the merged reads

were < 90 bp in length. After trimming sequences to 150 bp, sequences were quality filtered to a maximum expected error threshold of 1.0, which retained 99.7% of reads. Phylotypes were processed as exact sequence variants and then raw reads were mapped to a *de novo* database with UNOISE3 (Edgar and Valencia, 2018). Singletons (e.g. phylotypes represented by only one read) were excluded, and we additionally removed highly divergent sequences by filtering against the Barcode of Life (BOLD) database (Ratnasingham and Hebert, 2007), excluding phylotypes with less than an 85% identity match to any sequence in this reference database. Approximately 59% of the raw reads mapped to the *de novo* database, with 2.2 million sequence reads were retained. After this filtering process, the median number of total CO1 sequences retained per sample was 5631. The BOLD database was then used for taxonomic classification of 985 phylotypes (unique sequence variants) with the remaining 245 phylotypes that were unassigned against BOLD classified against the NCBI Genbank database using the BLAST algorithm. Lastly, all phylotypes not classified to at least the family level of taxonomic resolution were removed prior to statistical analyses.

2.4. Statistical analyses

All statistical analyses were executed in the R environment (version 3.4.2, R Core Team, 2013). We pooled extraction replicates and within-site samples to avoid pseudo-replication (e.g. multiple samples within a site treated as independent observations; Hurlbert, 1984). Samples were rarefied to 2000 reads per sample and 11 of 120 plot-level samples were discarded as they were below the 2000 read threshold for sequencing coverage. To standardize analyses across methods, we dropped plot-level samples across all methods if they were below the minimum threshold for one, which resulted in 87 plot-level samples retained and pooled at the site level. A family was considered present in a sample if there were at least three reads per sample from that family. Taxa were retained if detected in at least two samples for all methods. Finally, data were transformed into presence/absence as per Madden et al. (2016), as arthropod body size will bias the relative abundance estimates (e.g. sequence abundances are not proportional to biomass; Elbrecht and Leese, 2015).

To assess how the methods differed with respect to their estimates of arthropod richness, we ran pairwise Kruskal-Wallis rank sum tests. We then used Spearman's correlations to test whether richness (number of families per site) was correlated across the three methods. To investigate how similar the overall arthropod community composition was between the three methods across the 10 sites, we computed pairwise Jaccard distances from the presence-absence matrices and then calculated the correlation between all identification methods (morphology, arthropod barcode, and bulk soil barcode) using Mantel tests (Spearman correlation coefficient with 5000 permutations) in the vegan package (v2.4.1; Oksanen et al., 2016). All data are publicly available in the Figshare digital repository (http://dx.doi.org/10.6084/m9.figshare.5987635).

3. Results

3.1. Detection of soil arthropods across identification methods

We detected 19 arthropod orders in total across the three methods we used to characterize soil arthropod communities (Fig. 1): morphological identification of extracted arthropods, CO1 sequencing of DNA from the extracted arthropods, and CO1 sequencing of DNA extracted from bulk soil. Of the 19 orders detected, 16 were detected via morphological identification, 14 via arthropod DNA sequencing, and 12 via bulk soil DNA sequencing. Generally, the arthropod orders that were more ubiquitous across sites were detected with all three methods, including the trombidiform and sarcoptiform mites, the collembolan orders Entomobryomorpha, Poduromorpha, and Symphypleona, and the insect orders Coleoptera, and Diptera, with the exception of Lepidoptera which was frequently detected via arthropod DNA and bulk soil DNA analyses, but not via morphology (Fig. S1). In contrast, the less ubiquitous arthropod orders (e.g. those found only at one or two of the ten sites) were more variably detected across all three methods. For example, both Diplopoda orders were only detected via morphology or arthropod DNA and not recovered from bulk soil DNA.

At the family level, 90 families were recovered in total with 50 families detected via morphological identification, 60 via arthropod DNA sequencing, and 51 from bulk soil DNA sequencing (Fig. 2 and Fig. S2). Within the class Insecta (Fig. 2a), a majority of the coleopteran families were recovered by each of the three methods. However, for both Diptera and Lepidoptera, a greater number of families were detected by arthropod and bulk soil DNA sequencing in comparison to morphological identification. In contrast, within the class Arachnida, morphological identification consistently recovered more families than either sequence-based method within the mite orders Mesostigmata, Sarcoptiformes, and Trombidiformes (Fig. 2b). Collembola was the third most family-rich class (Fig. 2c) and a majority of families (at least 5 of 8) were detected by all of the identification methods.

3.2. Congruent estimates of arthropod community richness and composition

Despite some differences in the detection of individual groups (Figs. 1 and 2), the total richness of soil arthropods detected across the sites was similar regardless of the method employed. We found no significant differences in mean family richness across the three methods (P-value > 0.05; Kruskal-Wallis rank sum test). Further, richness across sites was strongly correlated across all methods (Fig. 3). Bulk soil DNA and arthropod DNA richness were most strongly correlated (Spearman's $\rho = 0.76$, $p < 0.01$) although morph ID richness was also correlated with both arthropod DNA richness and soil DNA richness ($\rho = 0.64$ with $P < 0.05$ and $\rho = 0.63$ with $P < 0.05$, respectively).

The differences in arthropod community composition across the 10 sites were also significantly correlated across the three methods (Fig. S3). The strongest correlation in composition was between morphology and arthropod DNA (Mantel $\rho = 0.68$; $P < 0.001$). The morphology versus bulk soil DNA and the arthropod versus bulk soil DNA comparisons were also well-correlated ($\rho = 0.63$; $P < 0.001$ and $\rho = 0.45$; $P < 0.01$, respectively). These results indicate that the three methods were similar with respect to their ability to quantify differences in overall arthropod community composition across the 10 sites.

4. Discussion

The goal of this study was to evaluate the utility of two methods of DNA metabarcoding for the study of soil arthropod communities across a range of soil types, comparing the metabarcoding results to more standard morphological analyses of extracted arthropods. Despite some site-level discrepancies across the three methods, we found a strong correspondence between the morphological and metabarcoding approaches in terms of their ability to detect taxonomic groups, overall richness observed, and overall community composition. We found that the differences across the sites were well-correlated across the three methods in terms of both richness (Fig. 3, Spearman's $\rho \geq 0.63$ with $P < 0.05$ for all comparisons) and community composition of arthropods (Fig. S3, Mantel $\rho \geq 0.45$ with $P < 0.05$ for all comparisons).

4.1. Differences in detection between morphological identification and DNA barcoding methods

The detection discrepancies between the morphological identification and the two sequence-based methods may be due to morphological incongruence, PCR and primer biases, or the presence of relic DNA in soil. First, morphological identity may not correspond directly to the taxonomic identity inferred from CO1 gene sequencing, a known

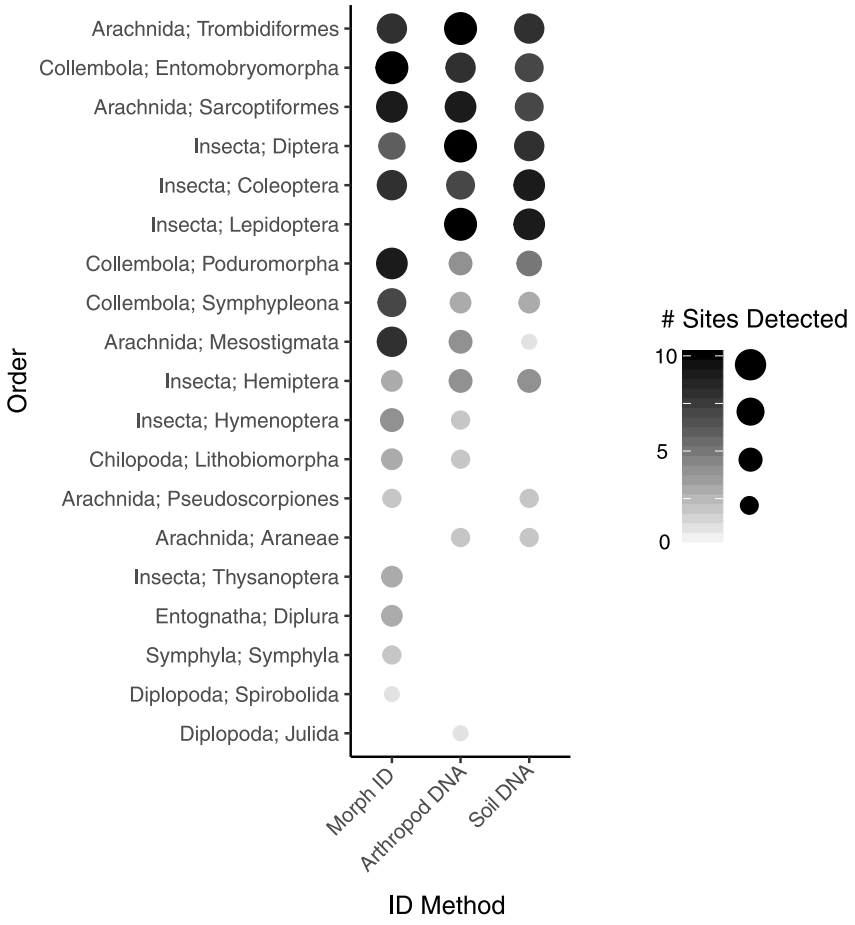

Fig. 1. Arthropod orders detected across the three identification methods including: morphological identification of funnel-extracted arthropods, CO1 gene sequencing of funnel-extracted arthropods, and CO1 sequencing of DNA extracted directly from bulk soil samples, referred to as 'morph ID', 'arthropod DNA', and 'bulk soil DNA'. Circles indicate that the order was detected for a given method. The circles are scaled in size from small to large and in gradient from white to black by occurrence (e.g. not detected = no circle and detected at 10 sites = large black circle). Arthropod orders are arranged by ubiquity across sites from top (most ubiquitous) to bottom (least). There were 19 orders detected in total with 16 detected from morphological identification, 14 from arthropod DNA, and 12 from bulk soil DNA.

challenge for some groups of arthropods (Wiemers and Fiedler, 2007) which can lead to incongruent taxonomic classifications. However, we would expect this to have less of an impact at the family level of resolution than at finer levels of taxonomic resolution. There was also likely incongruence introduced from those taxa that could not be confidently identified when using the different methods (such as specimens that were damaged and could not be confidently classified from morphology or, in the case of the DNA-based methods, CO1 sequence reads that had no database match at the family level). Second, there are known primer biases and differential amplification efficiencies across taxa (Clarke et al., 2014; Piñol et al., 2015) which may have influenced our results, even though we used a CO1 gene-targeting primer set that has been well-validated for arthropod analyses (Madden et al., 2016; Zeale et al., 2011). Nevertheless, primer and amplification bias, or lack of representation in reference databases, may explain why some families were detected via morphology and not with either of the CO1 gene sequencing approaches.

While amplification bias and lack of representation in reference database may explain methodological discrepancies for some smaller and poorly-studied arthropods, we were surprised that a few taxa which were common in the morphology-based identifications including the Formicidae (ants) and Tipulidae (crane flies) were not detected with either metabarcoding method. Although CO1 metabarcoding is used for ants (Smith et al., 2005; Smith and Fisher., 2009), the primers we used have known biases against Formicidae (Clarke et al., 2014) so this may explain why they were commonly detected via morphological analyses but infrequently detected when the CO1 gene sequencing methods were employed.

Third, we know that relic DNA (e.g. DNA from dead cells or extracellular DNA) can be common in soils (Carini et al., 2017). There is likely relic DNA from arthropods that are no longer present in the soil sample (e.g. arthropod DNA from shed cells or feces). While this may be more of an issue for the bulk soil DNA than for the extracted arthropod DNA, relic DNA detection likely impacts both methods as the extracted arthropods were not soil-free prior to DNA extraction. For example, the presence of relic DNA may explain why Lepidoptera were frequently detected via CO1 metabarcoding, but not at all with the morphological identification (Figs. 1 and 2). Similar results were obtained in a study of household dust where Lepidoptera were far more frequently detected in DNA-based surveys in comparison to morphology-based surveys (Madden et al., 2016), a pattern which they attributed to the presence of relic DNA in wing scales. Indeed, Lepidoptera wing scales preserve so well in soil and sediment habitats that they have been used to reconstruct paleoecological environments (Montoro Girona et al., 2018; Navarro et al., 2017; van Eldijk et al., 2018). Alternatively, detection of Lepidoptera could also be from the gut contents or feces of arthropods (such as spiders) that preyed on Lepidoptera larvae or adults (Symondson, 2002). Finally, there are PCR-amplification biases towards Lepidoptera that may contribute to the observed differences between morphological and molecular methods (Clarke et al., 2014).

4.2. Detection discrepancies between the two DNA barcoding approaches

We also observed some differences in community structure when we compared the two sequence-based barcoding methods (arthropod DNA versus bulk soil DNA). It is likely that some taxa were not detected in the bulk soil sample due to the small amount of bulk soil (less than a gram in total) used for DNA extractions. Indeed, we found that observed richness (number of families detected) as determined via metabarcoding of extracted arthropods was slightly higher at eight of ten sites (Fig. 3) than metabarcoding analyses of the relatively small amounts of bulk soil from which DNA was extracted. This finding

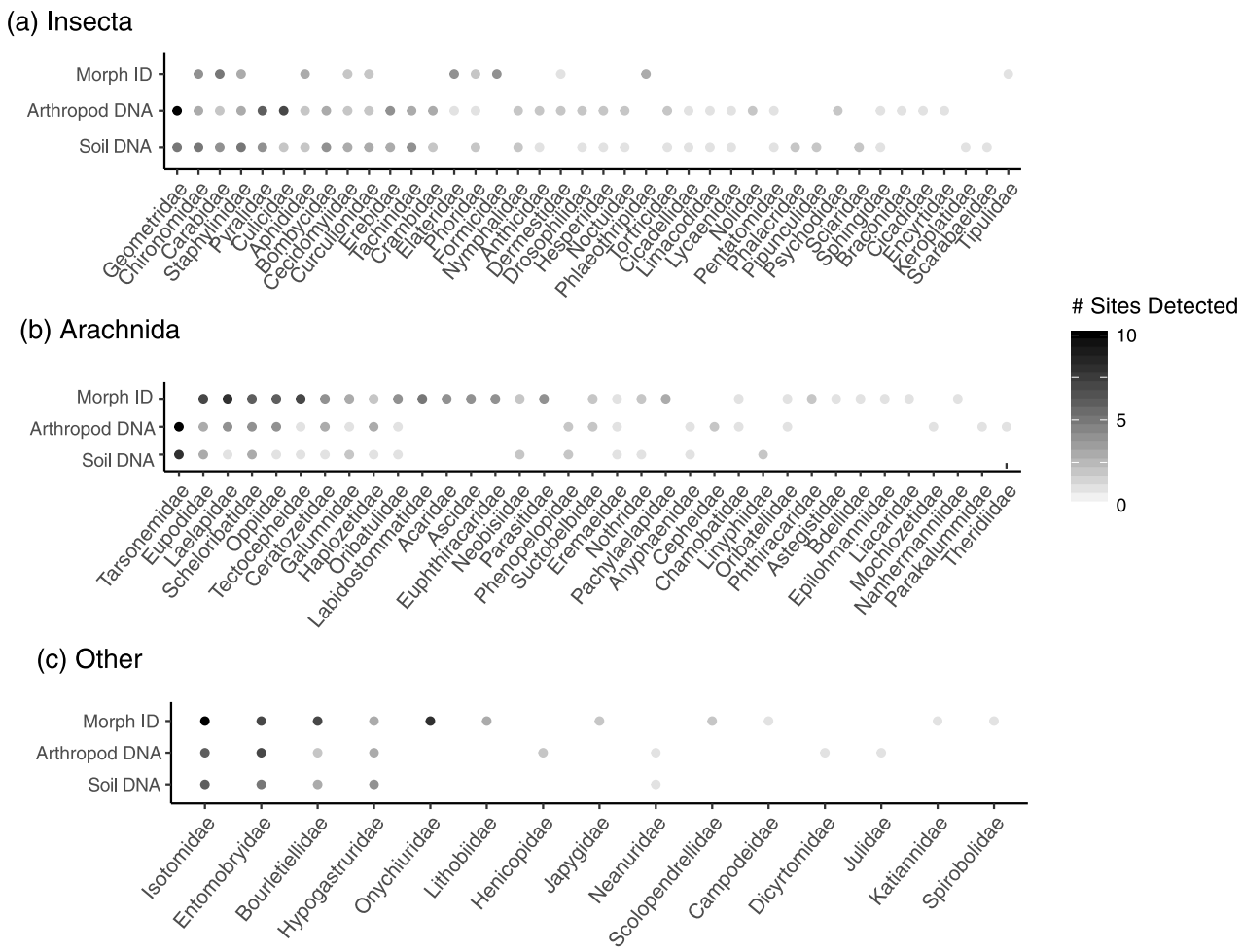

Fig. 2. Number of arthropod families detected across the three identification methods. Circles indicate families detected, and black to white gradient is scaled by occurrence across sites (e.g. white indicates that the family was not detected and black indicates detection at all 10 sites. For a-c, arthropod families are arranged along the x-axis by ubiquity across sites from left (most ubiquitous) to right (least). There were 90 families detected in total with 50 detected from morphological identification, 60 from arthropod DNA, and 51 from bulk soil DNA.

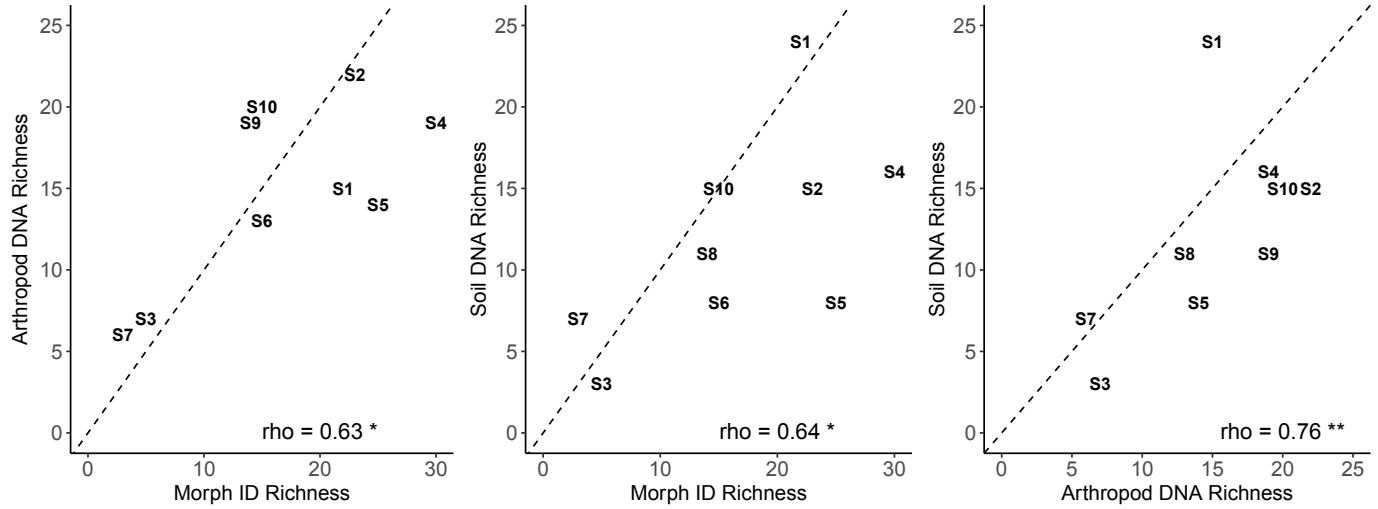

Fig. 3. Comparisons of richness at each site across methods. Spearman's correlations were used to test whether the three methods yield similar estimates of arthropod richness across the sites. Spearman's ρ and P-values provided for all below. Dashed line is a 1:1 line (e.g. if an equal number of taxa were identified across the two methods).

highlights the importance of extracting DNA from multiple sub-samples per soil or per site to detect arthropods which are often patchy in their spatial distributions (Ettema and Wardle, 2002; Maudsley et al., 2002).

4.3. Caveats and best practices for CO1 barcoding of soil arthropods

No specific approach provides an unbiased characterization of soil arthropod communities and we suggest that CO1 metabarcoding may be particularly useful for certain types of research questions, given the observed congruence with the morphological identification methods. For example, metabarcoding approaches enable more direct comparisons across datasets as the taxonomic identifications are not subject to differences in the taxonomic expertise of the individuals doing the analyses (Packer et al., 2018). Also, DNA sequence information can be more easily incorporated into multi-domain studies of soil communities. For example, targeting soil arthropods along with bacterial, archaeal, fungal, or protist communities that are often surveyed using similar molecular approaches (Gibson et al., 2014; Morriën et al., 2017). Lastly, the time necessary to extract and identify individual soil arthropods from more than a few soil samples is much greater than the time needed to obtain CO1 gene sequence data from bulk soil.

Likewise, it is important to recognize that the molecular-based approaches also have some limitations compared to traditional approaches that may be important for some studies. First, the methods proposed here are limited to presence rather than abundance-based metrics which may be critical for some research questions. Second, there is no corresponding specimen collection which may be important to discerning life history traits or for archival purposes. Finally, as we noted previously, the CO1 barcode likely has some limitations in discriminating between individual taxa due to lack of representation in sequence databases and there are potential biases towards some particular arthropod families (such as Lepidoptera; Clarke et al., 2014). Indeed, metabarcoding approaches for profiling soil arthropod communities will continue to improve as more taxa are identified morphologically with corresponding reference sequences.

For this study, soils were collected from ten sites around Ithaca, New York that were relatively different from each other. Our sites spanned the range of intensively managed (golf course) to relatively undisturbed (undisturbed forest), with broadly different vegetation types (Supplementary Table 1). Thus, while our results highlight congruence between morphological and sequence-based methods (e.g. the diversity and composition of arthropod communities found at particular sites were similar regardless of method employed) – it remains unclear to what extent we would be able to discern differences in arthropod communities across highly similar sites (for example, if we only sampled within hardwood forest sites).

We also briefly note a few 'best practices' for the generation and interpretation of CO1 gene sequence data from soil arthropod communities. First, it is likely important to pool multiple samples per site for DNA analyses to get a representative picture of the arthropods at a site given their often-patchy distributions. Second, the molecular-based approaches are not a valid proxy for abundances or relative abundances due to body size differences and given that DNA can be shed in frass (Elbrecht and Leese, 2015). Thus, we recommend using presence-only data with a detection threshold (e.g. minimum number of reads in a sample sufficient to call a specific taxon 'present'). Third, it is important to recognize that detecting DNA from a given taxon does not mean that taxon is currently living in that sample, as relic DNA (e.g. DNA from shed lepidopteran wing scales) can persist in soil, gut contents, or feces from predators.

In conclusion, our results suggest that metabarcoding approaches from extracted arthropods or bulk soil can be used to survey soil arthropod communities yielding similar estimates of community composition and diversity. While traditional morphological-based assessments are clearly valuable and have their own set of advantages, we show that a CO1 gene sequencing-based approach for characterizing soil arthropod communities may be a useful approach for some soil ecological research. This is particularly true for those studies that require reasonably rapid assessment of arthropod communities with the advantage that, by directly comparing the CO1 sequence data across sites, arthropod diversity and community composition can be directly compared across studies in situations where morphology-based identifications may be inconsistent or simply unavailable.

Acknowledgements

We thank Annise Dobson, Maya Montoya-Pimolwatana, Matt Gebert, and Jessica Henley for assistance with soil collection, sample processing and laboratory work. This project was funded by an NSF Graduate Research Fellowship to AMO, a grant from the Smithsonian Tropical Research Institute to NF, and a grant from the US Department of Agriculture to KW.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.soilbio.2018.06.026.

References

André, H.M., Ducarme, X., Lebrun, P., 2002. Soil biodiversity: myth, reality or conning? Oikos 96, 3–24.
André, H., Ducarme, X., Anderson, J., Crossley Jr., D., Koehler, H., Paoletti, M., Walter, D., Lebrun, P., 2001. Skilled eyes are needed to go on studying the richness of the soil. Nature 409, 761.
Bates, S.T., Clemente, J.C., Flores, G.E., Walters, W.A., Parfrey, L.W., Knight, R., Fierer, N., 2013. Global biogeography of highly diverse protistan communities in soil. The ISME Journal 7, 652.
Bienert, F., De Danieli, S., Miquel, C., Coissac, E., Poillot, C., Brun, J.J., Taberlet, P., 2012. Tracking earthworm communities from soil DNA. Molecular Ecology 21, 2017–2030.
Buddle, C.M., 2010. Photographic key to the Pseudoscorpions of Canada and the adjacent USA. Canadian Journal of Arthropod Identification 1–76.
Carini, P., Marsden, P.J., Leff, J.W., Morgan, E.E., Strickland, M.S., Fierer, N., 2017. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. Nature Microbiology 2, 16242.
Christiansen, K.A., Janssens, F., 2010. Checklist of the collembola: key to the common surface dwelling collembola from north America. accessed 2018. https://www.collembola.org/key/bugguide.htm.
Clarke, L.J., Soubrier, J., Weyrich, L.S., Cooper, A., 2014. Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. Molecular Ecology Resources 14, 1160–1170.
Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D.J., Bardgett, R.D., Maestre, F.T., Singh, B.K., Fierer, N., 2018. A global atlas of the dominant bacteria found in soil. Science 359, 320–325.
Edgar, R.C., Valencia, A., 2018. Updating the 97% identity threshold for 16SRNA OTUs. Bioinformatics bty 113.
Edgar, R.C., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods 10, 996.
Edwards, C.A., 1991. The assessment of populations of soil-inhabiting invertebrates. Agriculture, Ecosystems & Environment 34, 145–176.
Elbrecht, V., Leese, F., 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. PLoS One 10, e0130324.
Ettema, C.H., Wardle, D.A., 2002. Spatial soil ecology. Trends in Ecology & Evolution 17, 177–183.
Fierer, N., Strickland, M.S., Liptzin, D., Bradford, M.A., Cleveland, C.C., 2009. Global patterns in belowground communities. Ecology Letters 12, 1238–1249.
Gan, H., Wickings, K., 2017. Soil ecological responses to pest management in golf turf vary with management intensity, pesticide identity, and application program. Agriculture. Ecosystems and Environment 246, 66–77.
Gergócs, V., Hufnagel, L., 2009. Application of oribatid mites as indicators: review. Applied Ecology and Environmental Research 7, 79–98.
Gibson, J., Shokralla, S., Porter, T.M., King, I., van Konynenburg, S., Janzen, D.H., Hallwachs, W., Hajibabaei, M., 2014. Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. Proceedings of the National Academy of Sciences 111, 8007–8012.
Griffiths, B.S., de Groot, G.A., Laros, I., Stone, D., Geisen, S., 2018. The need for standardisation: exemplified by a description of the diversity, community structure and ecological indices of soil nematodes. Ecological Indicators 87, 43–46.
Griffiths, B.S., Donn, S., Neilson, R., Daniell, T.J., 2006. Molecular sequencing and morphological analysis of a nematode community. Applied Soil Ecology 32, 325–337.
Gulvik, M.E., 2007. Mites (Acari) as indicators of soil biodiversity and land use monitoring: a review. Polish Journal of Ecology 55, 415–440.
Hamady, M., Walker, J.J., Harris, J.K., Gold, N.J., Knight, R., 2008. Error-correcting

barcoded primers for pyrosequencing hundreds of samples in multiplex. Nature Methods 5, 235–237.
Hamilton, H.C., Strickland, M.S., Wickings, K., Bradford, M.A., Fierer, N., 2009. Surveying soil faunal communities using a direct molecular approach. Soil Biology & Biochemistry 41, 1311–1314.
Hogg, I.D., Hebert, P.D., 2004. Biological identification of springtails (Hexapoda: collembola) from the Canadian Arctic, using mitochondrial DNA barcodes. Canadian Journal of Zoology 82, 749–754.
Hurlbert, S.H., 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54, 187–211.
Ji, Y., Ashton, L., Pedley, S.M., Edwards, D.P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P.M., Woodcock, P., Edwards, F.A., Larsen, T.H., Hsu, W.W., Benedick, S., Hamer, K.C., Wilcove, D.S., Bruce, C., Wang, X., Levi, T., Lott, M., Emerson, B.C., Yu, D.W., 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. Ecology Letters 16, 1245–1257.
Kaston, B.J., 1978. How to Know the Spiders, third ed. William C. Brown Co, Dubuque, Iowa.
Krantz, G.W., Walter, D.E. (Eds.), 2009. A Manual of Acarology, third ed. Texas Tech University Press, Lubbock, TX.
Leff JW, 2016. URL: https://github.com/leffj/helper-code-for-uparse [accessed Dec 2017].
Macfadyen, A., 1962. Soil arthropod sampling. Advances in Ecological Research 1, 1–34.
Madden, A.A., Barberán, A., Bertone, M.A., Menninger, H.L., Dunn, R.R., Fierer, N., 2016. The diversity of arthropods in homes across the United States as determined by environmental DNA analyses. Molecular Ecology 25, 6214–6224.
Maudsley, M., Seeley, B., Lewis, O., 2002. Spatial distribution patterns of predatory arthropods within an English hedgerow in early winter in relation to habitat variables. Agriculture, Ecosystems & Environment 89, 77–89.
Montoro Girona, M., Navarro, L., Morin, H., 2018. A secret hidden in the sediments: Lepidopteran scales. Frontiers in Ecology and Evolution 6, 2.
Morriën, E., Hannula, S.E., Snoek, L.B., Helmsing, N.R., Zweers, H., de Hollander, M., Soto, R.L., Bouffaud, M.-L., Buee, M., Dimmers, W., Duyts, H., Geisen, S., Girlanda, M., Griffiths, R.I., Jørgensen, H.-B., Jensen, J., Plassart, P., Redecker, D., Schmelz, R.M., Schmidt, O., Thomson, B.C., Tisserant, E., Uroz, S., Winding, A., Bailey, M.J., Bonkowski, M., Faber, J.H., Martin, F., Lemanceau, P., de Boer, W., vanVeen, J.A., van der Putten, W.H., 2017. Soil networks become more connected and take up more carbon as nature restoration progresses. Nature Communications 8, 14349.
Navarro, L., Harvey, A.É., Morin, H., 2017. Lepidoptera wing scales: a new paleoecological indicator for reconstructing spruce budworm abundance. Canadian Journal of Forest Research 999, 1–7.
Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., 2016. Vegan: community ecology package R package version 2.4-1. accessed Dec 2017. https://CRAN.R-project.org/package=vegan.
Oliverio, A.M., Bradford, M.A., Fierer, N., 2017. Identifying the microbial taxa that consistently respond to soil warming across time and space. Global Change Biology 23, 2117–2129.
Packer, L., Monckton, S.K., Onuferko, T.M., Ferrari, R.R., 2018. Validating taxonomic identifications in entomological research. Insect Conservation and Diversity 11, 1–12.
Piñol, J., Mir, G., Gomez-Polo, P., Agustí, N., 2015. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. Molecular Ecology Resources 15, 819–830.
Porco, D., Decaëns, T., Deharveng, L., James, S.W., Skarżyński, D., Erséus, C., et al., 2013. Biological invasions in soil: DNA barcoding as a monitoring tool in a multiple taxa survey targeting European earthworms and springtails in North America. Biological Invasions 15, 899–910.
Querner, P., Bruckner, A., 2010. Combining pitfall traps and soil samples to collect Collembola for site scale biodiversity assessments. Applied Soil Ecology 45, 293–297.
R Core Team, 2013. R: a Language and Environment for Statistical Computing.
Ratnasingham, S., Hebert, P.D.N., 2007. BARCODING: bold: the barcode of life data system. Molecular Ecology Notes 7, 355–364. http://www.barcodinglife.org.
Read, D.S., Sheppard, S.K., Bruford, M.W., Glen, D.M., Symondson, W.O.C., 2006. Molecular detection of predation by soil micro-arthropods on nematodes. Molecular Ecology 15, 1963–1972.
Ruf, A., 1998. A maturity index for predatory soil mites (Mesostigmata: gamasina) as an indicator of environmental impact of pollution on forest soils. Applied Soil Ecology 9, 447–452.
Schaefer, M., Schauermann, J., 1990. The soil fauna of beech forests: comparison between a mull and a moder soil. Pedobiologia 34, 299–314.
Smith, M.A., Fisher, B.L., 2009. Invasions, DNA barcodes, and rapid biodiversity assessment using ants of Mauritius. Frontiers in Zoology 6, 31.
Smith, M.A., Fisher, B.L., Hebert, P.D., 2005. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. Philosophical Transactions of the Royal Society B: Biological Science 360, 1825–1834.
Smith, M.A., Rodriguez, J.J., Whitfield, J.B., Deans, A.R., Janzen, D.H., Hallwachs, W., Herbert, P.D.N., 2008. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proceedings of the National Academy of Sciences of the United States of America 105, 12359–12364.
Socarrás, A., 2013. Soil mesofauna: biological indicator of soil quality. Pastos Forrajes 36, 14–21.
Symondson, W.O.C., 2002. Molecular identification of prey in predator diets. Molecular Ecology 11, 627–641.
Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., et al., 2014. Global diversity and geography of soil fungi. Science 346, 1256688.
Treonis, A.M., Unangst, S.K., Kepler, R.M., Buyer, J.S., Cavigelli, M.A., Mirsky, S.B., Maul, J.E., 2018. Characterization of soil nematode communities in three cropping systems through morphological and DNA metabarcoding approaches. Scientific Reports 8, 2004.
Triplehorn, C.A., Johnson, N.F., 2005. Borror and DeLong's Introduction to the Study of Insects, seventh ed. Thomson Brooks/Cole, Davis, CA.
van Eldijk, T.J., Wappler, T., Strother, P.K., van der Weijst, C.M., Rajaei, H., Visscher, H., van de Schootbrugge, B., 2018. A Triassic-Jurassic window into the evolution of Lepidoptera. Science Advances 4, 1701568.
Waite, I.S., O'Donnell, A.G., Harrison, A., Davies, J.T., Colvan, S.R., Ekschmitt, K., et al., 2003. Design and evaluation of nematode 18S rDNA primers for PCR and denaturing gradient gel electrophoresis (DGGE) of soil community DNA. Soil Biology & Biochemistry 35, 1165–1173.
Wall, D.H., Fitter, A.H., Paul, E.A., 2005. Developing New Perspectives from Advances in Soil Biodiversity Research. Biological Diversity and Function in Soils. Cambridge University Press, Cambridge, pp. 31–43.
Wickings, K., Grandy, A.S., 2013. Management intensity interacts with litter chemistry and climate to drive temporal patterns in arthropod communities during decomposition. Pedobiologia 56, 105–112.
Wiemers, M., Fiedler, K., 2007. Does the DNA barcoding gap exist? –A case study in blue butterflies (Lepidoptera: lycaenidae). Frontiers in Zoology 4, 8.
Wu, T., Ayres, E., Bardgett, R.D., Wall, D.H., Garey, J.R., 2011. Molecular study of worldwide distribution and diversity of soil animals. Proceedings of the National Academy of Sciences 108, 17720–17725.
Wu, T., Ayres, E., Li, G., Bardgett, R.D., Wall, D.H., Garey, J.R., 2009. Molecular profiling of soil animal diversity in natural ecosystems: incongruence of molecular and morphological results. Soil Biology & Biochemistry 41, 849–857.
Yang, C., Wang, X., Miller, J.A., de Blécourt, M., Ji, Y., Yang, C., et al., 2014. Using metabarcoding to ask if easily collected soil and leaf-litter samples can be used as a general biodiversity indicator. Ecological Indicators 46, 379–389.
Yu, D.W., Ji, Y., Emerson, B.C., Wang, X., Ye, C., Yang, C., Ding, Z., 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods in Ecology and Evolution 3, 613–623.
Zeale, M.R.K., Butlin, R.K., Barker, G.L.A., Lees, D.C., Jones, G., 2011. Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces: DNA Barcoding. Molecular Ecology Resources 11, 236–244.
Zhang, C.Z., 1998. Myriapoda. In: Yin (Ed.), Pictorial Keys to Soil Animals of China. Science Press, Beijing, pp. 569–585.