Suitability of PCR primers for characterizing invertebrate communities from soil and leaf litter
targeting metazoan 18S ribosomal or cytochrome oxidase I (COI) genes.

Dean J. Horton^{1,*}, Mark W. Kershner, Christopher B. Blackwood

Department of Biological Sciences, Kent State University, Kent, OH 44242, USA

¹ Present address: College of Science and Engineering, Central Michigan University, Mount
Pleasant, MI 48859, USA

*Corresponding author. E-mail address: horto2dj@cmich.edu

Abstract

Sequencing of environmental DNA (eDNA) for community analysis (i.e., eDNA metabarcoding) is already well established and applied in molecular microbial ecology. However, molecular methods for use in soil animal studies require further development before high-throughput sequencing can be considered a reliable technique in community ecology. To aid in this effort, we compare primers that target two frequently used genetic markers, mitochondrial cytochrome oxidase (COI) and ribosomal 18S genes, to determine their utility in broad soil animal eDNA studies. DNA was analyzed from individually identified invertebrates to test the efficiency of the primer sets in successfully targeting animal DNA and identification through sequencing. Primers were also tested for amplification of faunal genes from forest soil and leaf litter eDNA. Targeting the 18S gene resulted in the most successful amplification and correct identification of a wide range of individual invertebrate taxa, and was the most reliable primer set for use in eDNA analysis of invertebrate communities spanning several phyla. In contrast, the selected COI primers were inefficient in identifying a wide range of invertebrates, and amplified mostly bacterial sequences from eDNA.

Keywords: invertebrate, soil, leaf litter, environmental DNA, mitochondrial COI, ribosomal 18S

1. Introduction

Terrestrial invertebrates, particularly those associated with soil, commonly manifest high taxonomic and functional diversity [1,2], often playing an important role in structuring local food webs. As a result, it is important to uncover and accurately characterize their biodiversity to

fully understand the relative importance of individual taxa within the complex soil community. However, quantifying species richness in soil can be a daunting task, with 1-m² of temperate forest soil often containing more than 1,000 invertebrate species [3]. These taxa can display wide phylogenetic diversity, spanning multiple phyla [4] and sizes, ranging from microscopic (e.g., nematodes and mites) to macroscopic (e.g., earthworms, centipedes and millipedes).

Traditionally, soil animal community studies have used identification methods based strictly on organismal morphology [5–8]. This approach requires significant taxonomic expertise, often restricting the number of taxa that can be simultaneously surveyed and deterring many scientists from studying soil invertebrate communities. Morphological identification approaches are also exceedingly time-consuming for taxonomists seeking to uncover biodiversity at lower taxonomic levels in large numbers of samples. Given the time requirements and importance of taxonomic expertise, robust molecular techniques that can be used to examine communities through mixed-species environmental DNA (eDNA) analysis may prove useful for unearthing soil biodiversity. These methods have the potential to identify a diverse community of invertebrates in large numbers of soil samples in a timely fashion [9]. In particular, high-throughput sequencing of eDNA is a promising method for rapidly revealing biodiversity in an increasingly inexpensive manner [10,11].

However, before high-throughput sequencing can become an effective tool for use in soil animal biodiversity studies, PCR primers targeting universal genetic markers must be validated [11]. Primer selection is a critical step in targeting specific taxonomic groups in eDNA analysis, as primers ultimately determine the accuracy and taxonomic breadth of acquired datasets. An optimal primer set should accurately amplify a targeted marker gene from eDNA, which can then be used to match sequences to specific, desired levels of taxonomic identification. Currently,

there are only two genes that can be targeted in invertebrate DNA studies with appropriate background databases for use in sequence identification. The mitochondrial cytochrome oxidase subunit I (COI) has been frequently used in DNA barcoding studies of individual invertebrate specimens in soil, above-ground, and aquatic environments [12–17], and can be effectively used to determine taxonomic identity at the species-level [18,19]. However, Yang et al. [20] targeted the COI gene within soil eDNA and found the selected primers to be inefficient for identification of eukaryotic taxa. The 18S ribosomal RNA gene has also been used in soil fauna surveys, including several eDNA studies [20–23]. However, invertebrate 18S ribosomal gene sequences can typically only be identified to higher taxonomic levels (i.e., not to the level of species) [20,24]. Thus, based upon previous studies, it is not clear which gene is most appropriate to target in eDNA studies of animal communities.

In this study, we sought to determine whether primers targeting 18S rRNA or COI genes would be suitable for targeting animal taxa in studies of eDNA obtained from soil and leaf litter. Primer sets were chosen based on their potential ability to target the widest possible phylogenetic range of animal DNA within eDNA samples. First, we tested these primers on genomic DNA extracted from individual voucher specimens of common temperate forest invertebrates to determine primer set success and taxonomic specificity of identification using the resulting genetic sequences. Second, using eDNA extracted from soil and leaf samples collected from a temperate forest, we determined if the chosen primer pairs were effective at uncovering broad invertebrate diversity in natural communities. Sanger sequencing was used for taxonomic identification of samples, as this method provided a more cost-effective and logistically feasible alternative to HTS at this stage of method development. We hypothesized that targeting the COI gene would result in invertebrate sequence identification at lower taxonomic levels (i.e., species-

level). However, given the results of previous eDNA studies [21,23], we also anticipated that 18S ribosomal sequences would allow for identification of a wider range of invertebrate taxa from eDNA. We also anticipated that targeting the 18S gene with the chosen primer set would result in identification of sequences at higher taxonomic levels (e.g., Class, Order, and Family).

2. Methods

2.1 Individual voucher specimen collection

Invertebrate specimens were collected from Jennings Woods (41.1723 N, 81.2022 W), a temperate forest in northeastern Ohio, USA (site description in Blackwood et al. [25]).

Invertebrates were obtained from pitfall traps that were opened on 15 September 2009 and filled with 4 cm of 95% ethanol and collected on 19 September 2009. To test primer pair breadth and specificity, 94 invertebrate taxa (stored in ethanol) were chosen, spanning multiple phyla. Each specimen was identified to varying levels of taxonomic resolution using microscopy (Table 1).

2.2 Primer testing on invertebrate voucher specimens

DNA was extracted from each specimen using the prepGEM™ Insect kit (ZyGEM, Hamilton, New Zealand). We destructively sampled whole micro- and mesofauna specimens (e.g., nematodes, collembolans, and mites) for DNA extraction, and used partial macrofauna specimens (e.g., leg of a camel cricket) so that the majority of the voucher could be retained. PCRs were completed with a Stratagene Robocycler thermocycler (Agilent Technologies, Santa Clara, CA). PCRs targeting a 658 bp region of the COI gene were carried out using primer set Fol-degen-for (5' TCNACNAAYCAYAARRAYATYGG 3') and Fol-degen-rev (5' TANACYTCNGGRTGNCCRAARAAYCA 3') [26]. These primers were selected because they

105 were designed to cover all Insecta sequences [26], and were the broadest primers we found
106 available for targeting COI. Additionally, these degenerate primers have not been previously
107 tested on bulk environmental samples. For optimization in PCR conditions for each sample, the
108 number of cycles and MgCl₂ and primer concentrations for each sample was varied to achieve
109 successful PCR amplification, as determined by gel electrophoresis. These optimization steps
110 were taken to ensure that PCR was tested exhaustively on each sample to test the ability for
111 selected primers to target and amplify DNA from voucher specimens. Failure of DNA
112 amplification during PCR was considered as a failure on the part of the primer set to amplify
113 DNA from that individual sample. During PCR optimization, final concentrations of reagents for
114 PCR reactions consisted of 0.03 U taq, 1X PCR buffer, 0.2 mM dNTPS, 1.5-2.5 mM MgCl₂, 1
115 mg/mL BSA, 0.16-0.4 μ M of each primer. Ranges of specific reagents during PCR reflected
116 different PCR optimization conditions found to be necessary for separate voucher specimens.
117 Conditions for the PCRs consisted of: initial denaturation at 94°C for 1 min., 35-50 cycles
118 (denaturing at 45 sec. at 94°C, annealing for 45 sec. at 40°C, and extension for 45 sec. at 72°C),
119 followed by a final extension at 72°C for 10 min. PCRs targeting an 830 bp region of the 18S
120 rRNA gene were run using primer set 18S11b (5' GTCAGAGGTTCGAAGGCG 3') and 18S2a
121 (5' GATCCTTCCGCAGGTTTACC 3') [21]. These primers are also expected to cover a broad
122 range of Metazoa and should therefore be useful in studies using eDNA [21]. Concentrations of
123 reagents for these reactions were the same as those described above. Conditions for these PCRs
124 consisted of: initial denaturation at 94°C for 3 min., 35-45 cycles (denaturation at 94°C for 30
125 sec., annealing at 55°C for 30 sec., and extension at 72°C for 90 sec.), followed by a final
126 extension at 72°C for 10 min. Each sample was replicated at optimal PCR conditions for each
127 specimen until enough amplicon DNA (COI and/or 18S) was obtained for sequencing. 18S PCRs

were purified using the UltraClean® PCR Clean-Up Kit (MoBio, Carlsbad, CA, USA). COI
PCRs were gel-purified using the UltraClean® GelSpin® DNA Extraction Kit (MoBio,
Carlsbad, CA, USA) due to the occurrence of probable nonspecific amplification of alternate
DNA segments. DNA concentration in cleaned-up PCRs was quantified using Quant-iT™
PicoGreen® dsDNA Reagent (Life Technologies, Carlsbad, CA, USA). All samples derived
from individual invertebrate specimens were sent to the Advanced Genetic Technologies Center
at the University of Kentucky, KY, USA for Sanger sequencing. Sequences were trimmed to
remove primer sequence and low-quality ends using Sequencher default settings (Gene Codes
Corporation, Ann Arbor, MI, USA). Sequences were then compared to the NCBI database using
the BLAST algorithm [27]. BLAST results were summarized using the last common ancestor
algorithm in MEGAN 5 [28] using default parameters. COI sequences were also tested against
the BOLD database (<http://www.boldsystems.org>). Sequences for which morphological and
molecular identification matched have been deposited in GenBank under accession numbers
KY554593 - KY554651.

2.3 eDNA in Environmental Samples

Soil and leaf litter samples were collected during August 2011 from Manistee National
Forest (44.25 N, 85.78 W), a temperate deciduous forest located in the northwestern region of
the Lower Peninsula of Michigan. In an attempt to target eDNA originating from a wide breadth
of terrestrial faunal species, three soil samples (10 cm deep/1.5 cm wide cores) and three leaf
litter samples (gathered from directly above the soil samples) were collected from varying
locations. Samples were collected with sterilized equipment and immediately placed on dry ice
for transportation back to the laboratory. In the lab, soil samples were passed through a 2mm

sieve. Leaves were homogenized by grinding in a ZM100 grinder (Retsch, Haan, Germany; 12 tooth rotor, 0.5 mm mesh ring sieve). Samples were stored at -80°C until DNA extraction.

eDNA was extracted from 0.25 g of each soil sample using the PowerSoil® DNA Isolation Kit (MoBio, Carlsbad, CA, USA). eDNA was extracted from 0.1 g of each leaf sample using the CTAB method [29]. PCRs were carried out using a Stratagene Robocycler (Agilent Technologies, Santa Clara, CA) thermocycler, using a similar approach as described above for the individual invertebrate voucher specimens. To obtain successful amplification, the annealing temperature needed to be adjusted to 40-47°C for COI primers to achieve amplification from eDNA. Successful PCRs were replicated and pooled together for quantification using Quant-iT™ PicoGreen® dsDNA Reagent (Life Technologies, Carlsbad, CA, USA). Using a 1% agarose gel, DNA cleanup was completed using GelSpin® DNA Extraction Kit (MoBio, Carlsbad, CA, USA) for DNA obtained from both primer sets to exclude bands that did not fall within the targeted range of amplicon sizes.

PCR amplicons from eDNA were cloned using a TA Cloning® Kit and the vector pCR 2.1 (Life Technologies, Carlsbad, CA, USA). Following ligation, plasmids were used to transform One Shot® TOP10 Chemically Competent *E. coli* cells (Life Technologies). Two-hundred eighty-eight colonies were picked from the six samples and grown in LB broth. Sequencing of clones was completed with M13 primers at the Advanced Genetic Technologies Center (University of Kentucky, KY, USA).

Sequences acquired from eDNA samples were analyzed for community composition using MEGAN 5 with LCA parameters similar to those used for individual invertebrate sample sequences described above. COI sequence data was also tested against the BOLD database. Sequence data was combined from soil and leaf litter samples for further analyses, as cloning

success from soil samples was low and sampling was uneven across samples. Environmental sequences have been deposited in GenBank under accession numbers KY554483 - KY554592.

3. Results

3.1 Invertebrate voucher specimens

Successful amplification of correctly sized PCR product occurred across a broad taxonomic range of arthropods, as well as Nematoda and Clitellata, for both targeted genes. Overall, 85 out of 94 (90.4%) individual invertebrate DNA samples were successfully amplified when targeting the 18S region during PCR. The COI gene was successfully amplified from only 58 out of 94 (61.7%) invertebrate DNA samples. Collembola (the taxonomic group with the greatest number of specimens tested, $n = 51$) amplified successfully using both primer sets, although the 18S primer set amplified DNA from 47 individuals while the COI primer set amplified DNA from 33 individuals. Both nematode specimens were successfully amplified by the 18S primer set, while the COI primer set failed to amplify this taxonomic group.

We obtained 18S amplicon sequences from 56 specimens, and COI amplicon sequences from 19 specimens, which represented a wide range of the successfully amplified taxa. Of the resulting 18S sequences, 51 out of 56 (91.1%) agreed with the corresponding morphological identifications (Table 1). In contrast, 14 of 19 (73.7%) COI sequences agreed with morphological identifications (Table 1). COI sequences did not match to any sequences within the BOLD database. 18S sequences could be classified to taxonomic levels between subclass and family, while correctly identified COI sequences ranged from superphylum to species (Table 1).

3.2 Environmental soil and leaf samples

For 18S sequences, 111 out of 132 (84.1%) of environmental sequences were identified as Metazoa across soil and leaf litter samples (Fig. 1). Soil and leaf litter data were pooled together for further analysis to focus on the comparison between primers and because of large differences in the growth of clones and numbers of sequences obtained (16 soil sequences versus 116 leaf litter sequences). Metazoan sequences were assigned to phyla including (but not limited to) Arthropoda, Tardigrada, Gastrotricha, and Nematoda, which demonstrates the wide range of taxa that were successfully amplified and identified from eDNA (Figs. 1 and 2). Despite the relatively small number of sequences obtained, they still spanned several taxonomic groups within Arthropoda and Nematoda (data not shown). Non-metazoan sequences obtained from environmental samples (21 out of 132) matched plants ($n = 3$), were assigned to taxonomic levels higher than Metazoa (such as “cellular organism” or “Eukaryota”) ($n = 5$), or identified as vector DNA ($n = 13$). Taxonomic specificity of 18S sequence classifications ranged from phylum down to genus, depending on the taxa (Fig. 2). Results are more representative of the leaf litter PCR amplicons, instead of soil, due to more successful cloning reactions for leaf litter.

In contrast, only 8 out of 138 (5.8%) environmental COI sequences were identified as belonging to Metazoa across leaf litter and soil samples (Fig. 1). Most COI sequences were bacterial ($n = 75$), assigned to taxonomic levels greater/other than Metazoa ($n = 6$), or did not match with sequences within the NCBI database ($n = 47$). Of the eight metazoan sequences uncovered, few of these sequences had their best matches to sequences from animals that could be present in a temperate forest ecosystem. COI sequences tested against the BOLD database also predominantly identified as bacterial sequences, while other sequences did not match to a sequence within the database.

4. Discussion

Environmental DNA-sequencing of invertebrate communities based on metabarcoding depends on use of efficient and accurate PCR primers that target an informative gene shared by this broad range of organisms. Our results demonstrate high performance of selected 18S primers, as compared with an assay targeting the COI gene with degenerate primers. The 18S primers amplified a broader taxonomic range of invertebrates from both voucher specimens and eDNA, and were more specific to invertebrates in eDNA when compared with the COI primers, despite COI being suggested as a favorable alternative to the 18S marker [24]. However, it should be recognized that only one degenerate primer set was tested targeting COI, so further development and testing of alternate primers could eliminate this issue. Indeed, it is likely necessary that degenerate primers targeting the COI must be developed for inclusion of a wider breadth of invertebrate taxa, as has recently been accomplished for freshwater invertebrates [30].

COI sequences acquired from mixed-species eDNA were largely of bacterial origin (Fig. 2), a result consistent with Yang et al. [20]. The presence of bacterial DNA in eDNA was not surprising as eDNA contains a diverse mixture of DNA originating from organisms spanning the three domains. Thus, a proper primer set for eDNA studies must be able to discriminate against all DNA sequences of non-target organisms present in a sample (in this case, bacteria, fungi, plants, etc.) [31]. For example, *Wolbachia* nonspecific amplification can occur using a variety of primers that target the COI gene, specifically within insects [32]. Our results suggest that the flanking regions commonly used as targets for PCR amplification of metazoan COI genes are similar to regions of DNA within genomes of other taxonomic groups, including the bacteria amplified from eDNA in this study. The degeneracy of the COI primer set may partly explain why bacterial sequences were predominantly amplified during PCR [26], although it is unlikely

that this attribute alone accounted for all issues. Despite non-specificity possibly stemming from degeneracies in the COI primers, these degeneracies are likely necessary for COI primers to be able to target a broad range of invertebrate community members due to the less conserved nature of the COI gene amongst Metazoa. Although degenerate COI primers appear to have limited use in soil eDNA, they may still be useful in studies seeking to identify individual specimens [12-19] or survey mixtures of invertebrates that have been extracted from various sampling approaches and then homogenized into a single sample [26,33]. These approaches help exclude organisms from alternate taxonomic groups found within the soil (e.g., bacteria, fungi, plants, etc.) prior to DNA extraction. However, many invertebrate taxa are difficult to isolate by these extraction methods, suggesting that extraction of eDNA directly from the soil is a useful approach to comprehensively target local animal fauna. Genetic material from COI was also difficult to amplify within voucher specimens, and it has been previously observed that the Folmer primer set may not be as universal as once presumed [34,35]. A final problem noted with the primer set chosen to target the COI gene was that some COI sequences of metazoan descent matched database sequences of organisms unlikely to be present in our terrestrial samples, including marine or freshwater taxa. However, despite the problems noted with primer specificity, of the 19 COI sequences that were obtained from invertebrate specimens, 9 of them were identified to the level of genus and/or species (Table 1).

The 18S primer set appears to have significant advantages for use in eDNA studies targeting invertebrates spanning a wide breadth of phyla due to its ability to specifically and consistently amplify animal DNA, despite limitations in classifying sequences to lower taxonomic levels [20]. This problem may not be as extensive as previously indicated, however,

because we found that sequences could be confidently identified to taxonomic levels as low as family or genus in several taxa (Table 1).

Although targeting environmental DNA has been proven to be useful for biodiversity studies, caution must be taken when interpreting obtained sequence data. For example, degradation and persistence of eDNA depends on specific local environmental conditions [36] and eDNA could last for many years within an environment, which could in turn originate from both living or dead organisms [37]. Additionally, gene copy numbers vary between species [21], and PCR biases can preferentially amplify the DNA of organisms dependent upon annealing affinity of the primer set [37]. Taxonomic identification through eDNA sequencing is also dependent upon the specificity and conservation of the gene being targeted and the reference database for that gene [11], which is reliant on previous sequencing of morphologically identified taxa. Sequences generated within our study were tested against NCBI and BOLD databases, which are two of the largest global databases of invertebrate DNA sequences for the 18S and COI genes. Nevertheless, with the above caveats, data obtained from eDNA sequencing must be carefully interpreted.

5. Conclusions

This study clearly suggests that the primer set used to target the 18S ribosomal gene is favorable to the degenerate primer set used to target the COI gene. Until other COI primers are developed and validated, we recommend using 18S primers in studies targeting community faunal DNA from terrestrial environmental samples. While the COI marker has been found useful in identification of metazoans to the species level in specimen-level barcoding studies, often using specific primer sets tailored to narrow taxonomic groups, it is clear that one of the

most degenerate primer sets currently available for the COI gene is not effective at separating animal DNA from bulk environmental DNA for community-level studies. In contrast, the accuracy and consistency of results generated by the 18S primers provide a reliable assay for eDNA studies centered on surveys of metazoan taxa.

6. Acknowledgements

This research was supported by grants from the US National Science Foundation (DEB-0918240 and DEB-0918878) and US Department of Energy (DE-SC0004335). Samples were obtained from Manistee National Forest with assistance from M. Gacura, S. Eisenlord and E. Ryee.

7. References

- [1] P. García-Palacios, F.T. Maestre, J. Kattge, D.H. Wall, Climate and litter quality differently modulate the effects of soil fauna on litter decomposition across biomes, *Ecol. Lett.* 16 (2013) 1045–1053. doi:10.1111/ele.12137.
- [2] D.A. Wardle, The influence of biotic interactions on soil biodiversity, *Ecol. Lett.* 9 (2006) 870–886. doi:10.1111/j.1461-0248.2006.00931.x.
- [3] T. Decaëns, Macroecological patterns in soil communities, *Glob. Ecol. Biogeogr.* 19 (2010) 287–302. doi:10.1111/j.1466-8238.2009.00517.x.
- [4] P.S. Giller, The diversity of soil communities, the “poor man”’s tropical rainforest’, *Biodivers. Conserv.* 5 (1996) 135–168. doi:10.1007/BF00055827.
- [5] A.J.E. Barrett, R.A. Virginia, D.H. Wall, A.N. Parsons, L.E. Powers, M.B. Burkins, Variation in Biogeochemistry and Soil Biodiversity across Spatial Scales in a Polar Desert Ecosystem, *Ecology.* 85 (2004) 3105–3118.

- 311 [6] R.A. Hansen, Effects of Habitat Complexity and Composition on a Diverse Litter
312 Microarthropod Assemblage, *Ecology*. 81 (2000) 1120–1132.
- 313 [7] J.-A. Salamon, J. Alpehi, A. Ruf, M. Schaefer, S. Scheu, K. Schneider, A. Sühlig, M.
314 Maraun, Transitory dynamic effects in the soil invertebrate community in a temperate
315 deciduous forest: Effects of resource quality, *Soil Biol. Biochem.* 38 (2006) 209–221.
316 doi:10.1016/j.soilbio.2005.04.033.
- 317 [8] A.M. Treonis, D.H. Wall, R.A. Virginia, Invertebrate biodiversity in antarctic dry valley
318 soils and sediments, *Ecosystems*. 2 (1999) 482–492.
- 319 [9] P. Taberlet, E. Coissac, M. Hajibabaei, L.H. Rieseberg, Environmental DNA, *Mol.*
320 (2012) 1789–1793. [http://onlinelibrary.wiley.com/doi/10.1111/j.1365-](http://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2012.05542.x/full%5Cnpapers2://publication/uuid/3E2B6116-3B64-435F-A8EF-EFD5323DFD46)
321 [294X.2012.05542.x/full%5Cnpapers2://publication/uuid/3E2B6116-3B64-435F-A8EF-](http://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2012.05542.x/full%5Cnpapers2://publication/uuid/3E2B6116-3B64-435F-A8EF-EFD5323DFD46)
322 [EFD5323DFD46](http://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2012.05542.x/full%5Cnpapers2://publication/uuid/3E2B6116-3B64-435F-A8EF-EFD5323DFD46).
- 323 [10] S. Shokralla, J.L. Spall, J.F. Gibson, M. Hajibabaei, Next-generation sequencing
324 technologies for environmental DNA research, *Mol. Ecol.* 21 (2012) 1794–1805.
325 doi:10.1111/j.1365-294X.2012.05538.x.
- 326 [11] P.P. Taberlet, E. Coissac, F.F. Pompanon, C. Brochmann, E. Willerslev, Towards next-
327 generation biodiversity assessment using DNA metabarcoding, *Mol. Ecol.* 21 (2012)
328 2045–2050. doi:10.1111/j.1365-294X.2012.05470.x.
- 329 [12] R.D.H. Barrett, P.D.N. Hebert, Identifying spiders through DNA barcodes, *Can. J. Zool.*
330 83 (2005) 481–491. doi:10.1139/Z05-024.
- 331 [13] J.J. Huang, Q. Xu, Z.J.Z.J. Sun, G.L.G.L. Tang, Z.Y.Z.Y. Su, Z. Jun, G. Lan, Z. You,
332 Identifying earthworms through DNA barcodes, *Pedobiologia (Jena)*. 51 (2007) 301–309.
333 doi:10.1016/j.pedobi.2007.05.003.

- 334 [14] R. Floyd, E. Abebe, A. Papert, M. Blaxter, Molecular barcodes for soil nematode
335 identification, *Mol. Ecol.* 11 (2002) 839–850. doi:10.1046/j.1365-294X.2002.01485.x.
- 336 [15] F.O. Costa, J.R. DeWaard, J. Boutillier, S. Ratnasingham, R.T. Dooh, M. Hajibabaei,
337 P.D.N. Hebert, Biological identifications through DNA barcodes: the case of the
338 Crustacea, *Can. J. Fish. Aquat. Sci.* 64 (2007) 272–295. doi:10.1139/f07-008.
- 339 [16] M. Hajibabaei, D.H. Janzen, J.M. Burns, W. Hallwachs, P.D.N. Hebert, DNA barcodes
340 distinguish species of tropical Lepidoptera., *Proc. Natl. Acad. Sci. U. S. A.* 103 (2006)
341 968–971. doi:10.1073/pnas.0510466103.
- 342 [17] S. Jung, R.K. Duwal, S. Lee, COI barcoding of true bugs (Insecta, Heteroptera), *Mol.*
343 *Ecol. Resour.* 11 (2011) 266–270. doi:10.1111/j.1755-0998.2010.02945.x.
- 344 [18] P.D.N. Hebert, S. Ratnasingham, J. Waard, Barcoding animal life : cytochrome c oxidase
345 subunit 1 divergences among closely related species Barcoding animal life : cytochrome c
346 oxidase subunit 1 divergences among closely related species, *Proc. R. Soc. Lond. B.* 270
347 (2003) S96–S99. doi:10.1098/rsbl.2003.0025.
- 348 [19] P.D.N. Hebert, A. Cywinska, S.L. Ball, J.R. deWaard, Biological identifications through
349 DNA barcodes., *Proc. Biol. Sci.* 270 (2003) 313–21. doi:10.1098/rspb.2002.2218.
- 350 [20] C.X. Yang, Y.Q. Ji, X.Y. Wang, C.Y. Yang, D.W. Yu, Testing three pipelines for 18S
351 rDNA-based metabarcoding of soil faunal diversity, *Sci. China Life Sci.* 56 (2013) 73–81.
352 doi:10.1007/s11427-012-4423-7.
- 353 [21] H.C. Hamilton, M.S. Strickland, K. Wickings, M.A. Bradford, N. Fierer, Surveying soil
354 faunal communities using a direct molecular approach, *Soil Biol. Biochem.* 41 (2009)
355 1311–1314. doi:10.1016/j.soilbio.2009.03.021.
- 356 [22] M.S. Robeson 2nd, E.K. Costello, K.R. Freeman, J. Whiting, B. Adams, A.P. Martin,

- S.K. Schmidt, Environmental DNA sequencing primers for eutardigrades and bdelloid rotifers, *BMC Ecol.* 9 (2009) 25. doi:10.1186/1472-6785-9-25.
- [23] T. Wu, E. Ayres, R.D. Bardgett, D.H. Wall, J.R. Garey, Molecular study of worldwide distribution and diversity of soil animals, *Proc. Natl. Acad. Sci.* 108 (2011) 17720–17725. doi:10.1073/pnas.1103824108.
- [24] C.Q. Tang, F. Leasi, U. Obertegger, A. Kieneker, T.G. Barraclough, D. Fontaneto, The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna, *Proc. Natl. Acad. Sci.* 109 (2012) 16208–16212. doi:10.1073/pnas.1209160109.
- [25] C.B. Blackwood, K.A. Smemo, M.W. Kershner, L.M. Feinstein, O.J. Valverde-Barrantes, Decay of ecosystem differences and decoupling of tree community-soil environment relationships at ecotones, *Ecol. Monogr.* 83 (2013) 403–417. doi:10.1890/12-1513.1.
- [26] D.W. Yu, Y. Ji, B.C. Emerson, X. Wang, C. Ye, C. Yang, Z. Ding, Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring, *Methods Ecol. Evol.* 3 (2012) 613–623. doi:10.1111/j.2041-210X.2012.00198.x.
- [27] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, *J. Mol. Biol.* 215 (1990) 403–410. doi:10.1016/s0022-2836(05)80360-2.
- [28] D. Huson, S. Mitra, H. Ruscheweyh, Integrative analysis of environmental sequences using MEGAN4, *Genome Res.* 21 (2011) 1552–1560. doi:10.1101/gr.120618.111.Freely.
- [29] L. Wu, L.M. Feinstein, O. Valverde-Barrantes, M.W. Kershner, L.G. Leff, C.B. Blackwood, Placing the Effects of Leaf Litter Diversity on Saprotrophic Microorganisms in the Context of Leaf Type and Habitat, *Microb. Ecol.* 61 (2011) 399–409. doi:10.1007/s00248-010-9760-0.

- 380 [30] V. Elbrecht, F. Leese, Validation and development of COI metabarcoding primers for
381 freshwater macroinvertebrate bioassessment, *Front. Environ. Sci.* 5 (2017).
382 doi:10.3389/fenvs.2017.00011.
- 383 [31] M.G.A. Van Der Heijden, R.D. Bardgett, N.M. Van Straalen, The unseen majority: Soil
384 microbes as drivers of plant diversity and productivity in terrestrial ecosystems, *Ecol. Lett.*
385 11 (2008) 296–310. doi:10.1111/j.1461-0248.2007.01139.x.
- 386 [32] M.A. Smith, C. Bertrand, K. Crosby, E.S. Eveleigh, J. Fernandez-Triana, B.L. Fisher, J.
387 Gibbs, M. Hajibabaei, W. Hallwachs, K. Hind, J. Hrcek, D.W. Huang, M. Janda, D.H.
388 Janzen, Y. Li, S.E. Miller, L. Packer, D. Quicke, S. Ratnasingham, J. Rodriguez, R.
389 Rougerie, M.R. Shaw, C. Sheffield, J.K. Stahlhut, D. Steinke, J. Whitfield, M. Wood, X.
390 Zhou, Wolbachia and DNA barcoding insects: Patterns, potential, and problems, *PLoS*
391 *One.* 7 (2012). doi:10.1371/journal.pone.0036514.
- 392 [33] P. Arribas, C. Andújar, K. Hopkins, M. Shepherd, A.P. Vogler, D. Yu, Metabarcoding and
393 mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil,
394 *Methods Ecol. Evol.* 7 (2016) 1071–1081. doi:10.1111/2041-210X.12557.
- 395 [34] J. Geller, C. Meyer, M. Parker, H. Hawk, Redesign of PCR primers for mitochondrial
396 cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic
397 surveys, *Mol. Ecol. Resour.* 13 (2013) 851–861. doi:10.1111/1755-0998.12138.
- 398 [35] J. Lobo, P.M. Costa, M. a. L. Teixeira, M.S.G. Ferreira, M.H. Costa, F.O. Costa,
399 Enhanced primers for amplification of DNA barcodes from a broad range of marine
400 metazoans., *BMC Ecol.* 13 (2013) 34. doi:10.1186/1472-6785-13-34.
- 401 [36] K. Andersen, K.L. Bird, M. Rasmussen, J. Haile, H. Breuning-Madsen, K.H. Kjær, L.
402 Orlando, M.T.P. Gilbert, E. Willerslev, Meta-barcoding of “dirt” DNA from soil reflects

403 vertebrate biodiversity, *Mol. Ecol.* 21 (2012) 1966–1979. doi:10.1111/j.1365-
404 294X.2011.05261.x.
405 [37] P.F. Thomsen, E. Willerslev, Environmental DNA - An emerging tool in conservation for
406 monitoring past and present biodiversity, *Biol. Conserv.* 183 (2015) 4–18.
407 doi:10.1016/j.biocon.2014.11.019.
408
409

Table 1. Comparison of invertebrate specimen identifications determined using traditional (morphological) and molecular (sequencing of 18S rRNA and mitochondrial COI genes) approaches. E-value represents Expect (E) value and Score represents Total Score for the top hit sequence result testing against the NCBI database using BLAST. Numbers adjacent to Collembola identifications correspond to the number of individuals identified to that taxonomic rank. For all other taxa, there was only a single specimen included in the morphological and molecular identifications. The taxonomic rank for each identification is listed in parentheses following the rank (Codes = P: Phylum; C: Class; sC: Subclass; O: Order; sO: Suborder; iO: Infraorder; SF: Superfamily; F: Family; sF: Subfamily; G: Genus; S: Species). The symbol ‘◇’ indicates taxonomic mismatch between morphological and sequence identifications. “No Hits” indicates lack of sequence match to sequences within the NCBI database. Blank cells represent a lack of sequence representation for that gene due to lack of PCR amplification success. Assigned taxonomic ranks for molecular identifications were acquired by BLAST against the NCBI taxonomic database, followed by last common ancestor analysis using MEGAN.

<u>Morphological ID</u>	<u>18S DNA ID</u>	<u>E-value</u>	<u>Score</u>	<u>COI DNA ID</u>	<u>E-value</u>	<u>Score</u>
Collembola - Bourlettiellidae (2; F)	Collembola (1; C)	0	1293.8	◇ - <i>Hypogastrura</i> (1; G)	8.30 e-55	224.6
	Symphyleona (1; sO)	0	1376.9			
Collembola - Dicyrtomidae (7; F)	Collembola (1; C)	0	1293.8	Arthropoda (1; P)	1.15 e-03	54.7
	Symphyleona (6; sO)	0	1286.4 - 1299.3	No Hits (1)	NA	NA
Collembola - Entomobryidae (18; F)	Entomobryidae (12; F)	0	1304.8 - 1351.0	Paronellidae (1; F)	5.81 e-126	
	Entomobryoidea (2; SF)	0	1057.4 - 1275.3	<i>Lepidocyrtus paradoxus</i> (1; S)	3.05 e-39	172.9

	Orchesellinae (4; sF)	0	1297.5 - 1365.8	<i>Entomobrya</i> (1; G)	8.79 e-15	
				No Hits (1)	NA	NA
Collembola - Hypogastruridae (1; F)				◇ - Entomobryidae (1; F)	1.45 e-22	117.5
Collembola - Isotomidae (3; F)	<i>Isotoma</i> (2; G)	0	1255.0 - 1258.7			
	Isotomidae (1; F)	0	1306.7			
Collembola - Katiannidae (2; F)	Symphyleona (2; sO)	0	1225.4 - 1363.9			
Collembola - Tomoceridae (6; F)	Tomoceridae (6; F)	0	1284.5 - 1304.8	Collembola (1; C)	1.12 e-18	104.5
Coleoptera (O)	Nitidulidae (F)	0	1310.4			
Coleoptera - <i>Pterostichus melanarius</i> (S)				◇ - <i>Pterostichus coracinus</i> (S)	0	821.0
Coleoptera -Curculionidae (F)	◇ - Psocidae (F)	NA	NA			
Coleoptera -Staphylinidae (F)	◇ - Psocidae (F)	NA	NA			
Diptera (O)	Muscomorpha (iO)	0	1460.0	Brachycera (sO)	3.26 e-173	617.9
Hemiptera (O)	Heteroptera (sO)	0	1231.0	<i>Drymus</i> (2; G)	0 - 1.51 e-176	629 - 653

Hemiptera - Cicadellidae (F)	Hemiptera (O)	0	1363.9			
Hymenoptera - <i>Camponotus pennsylvanicus</i> (S)				<i>Camponotus</i> (G)	4.96 e-57	232.0
Hymenoptera - Diapriidae (F)	Hymenoptera (O)	0	1382.4			
Lepidoptera (O)	Neolepidoptera (iO)	0	1423.0	<i>Renia adspergillus</i> (S)	3.52 e-118	435.1
Orthoptera - <i>Ceuthophilus maculatus</i> (S)	Rhaphidophoridae (F)	0	1406.4			
Psocodea (O)				Neoptera (sC)	3.73 e-78	302.1
Arachnida - Acari Morph 1 (sC)	<i>Anystis</i> (G)	0	1369.5			
Arachnida - Acari Morph 2 (sC)	Monogynaspida (sO)	0	1231.0			
Arachnida - <i>Wadotes hybridus</i> (S)	◊ - Psocidae (F)	0	1330.7			
Arachnida - Opiliones (O)	Palpatores (sO)	0	1365.8	<i>Leiobunum vittatum</i> (S)	3.89 e-48	202.4
Arachnida - Pseudoscorpionida (O)		NA	NA	<i>Chthonius tetrachelatus</i> (S)	9.73 e-124	453.5
Diplopoda - Julidae (F)	Helminthomorpha (sC)	0	1149.7			
Diplopoda - <i>Oxidus gracilis</i> (S)	Polydesmida (O)	0	1352.9			

421

Isopoda - <i>Oniscus asellus</i> (S)	◇ - Psocidae (F)	0	1334.4	Arthropoda (P)	1.68 e-96	363.1
Clitellata - Lumbricidae (F)	Megascolecidae (F)	0	1210.7			
Nematoda (P)	◇ - Psocidae (F)	0	856.1