Restoration Ecology

Technical Article

DNA metabarcoding as a new tool for invertebrate community monitoring – a case study comparison with conventional techniques

Corinne Watts[1], Andrew Dopheide[2] Robert Holdaway[3], Carina Davis[3], Jamie Wood[3], Danny Thornburrow[1] and Ian A. Dickie[4]

[1]Landcare Research, Private Bag 3127, Hamilton 3216, New Zealand

[2]Landcare Research, Private Bag 92170, Auckland 1142, New Zealand

[3]Landcare Research, PO Box 40, Lincoln 7640, New Zealand

[4]Bio-Protection Research Centre, School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand

Running Head Monitoring invertebrates using DNA metabarcoding

Keywords: environmental DNA, Cape to City, COI, edge effects, forest, restoration

Author contributions: CW and RH designed the research; CW and DT collected the data and identified the invertebrates; CD and JW performed the laboratory analysis; ID and AD ran the bioinformatics, AD analysed the data; CW, AD and RH wrote and edited the manuscript.

Abstract

When restoring native biodiversity, it is particularly important to consider invertebrates, a diverse and functionally important component of biodiversity. However, their inclusion in monitoring and conservation planning has lagged behind larger fauna because collecting, sorting, and identifying invertebrates using conventional monitoring techniques is often expensive, time-consuming, and restricted by expertise in diagnostics. Emerging DNA metabarcoding techniques could potentially revolutionise monitoring of invertebrates by providing the ability to characterise entire communities from a single, easily collected environmental sample. We aimed to characterise the invertebrate fauna of an isolated, coastal forest fragment in New Zealand using the same level of financial investment for conventional invertebrate monitoring and two DNA metabarcoding approaches (analysis of DNA extracted from bulk invertebrate samples from conventional pitfall and malaise traps, and DNA extracted from soil samples). The bulk invertebrate DNA metabarcoding method was able to reproduce ecological patterns observed in the beetle community detected using conventional sampling. The soil DNA metabarcoding method detected a different beetle community and a more diverse array of invertebrate taxa than conventional sampling techniques, providing additional biodiversity data at no extra cost. DNA metabarcoding offers restoration managers a practical, cost-effective technique for characterizing whole invertebrate communities. However, increasing the taxonomic coverage of reference sequence databases (particularly for New Zealand invertebrates) through DNA barcoding efforts would improve the utility of metabarcoding methods for invertebrate monitoring, and should be a focus of future research.

Implications for practice

- DNA metabarcoding methods detected a diverse array of invertebrate taxa – more than conventional pitfall and malaise sampling techniques for the same cost.
- Metabarcoding analysis of bulk invertebrate DNA was able to reproduce ecological patterns in beetle community composition detected using conventional sampling, whereas metabarcoding analysis of soil DNA provided comparable data on other groups at no extra cost.
- DNA metabarcoding is a powerful tool for efficiently characterizing and monitoring whole invertebrate communities, from either bulk invertebrate samples collected from traps, or from soil samples. However, the sampling method has very strong effects on the inferred taxonomic community composition, as each method recovers a somewhat different component of the ecosystem.
- A lack of DNA reference data, particularly for New Zealand invertebrates, means that additional DNA barcoding would greatly improve the utility of metabarcoding methods as a tool for monitoring invertebrate communities.

Introduction

Consideration of invertebrates in restoration and management of native biodiversity is critical as invertebrates constitute a substantial proportion of biodiversity and are critical to ecosystem functions (Forup & Memmott 2005; Lavelle et al. 2006; Watts et al. 2008; Gregoire Taillefer & Wheeler 2012). However, their inclusion in biodiversity monitoring and conservation planning has lagged behind better-known, more widely appreciated taxa (Gerlach et al. 2013). When invertebrates are included, they are usually limited to large iconic invertebrates such as carabid beetles in Europe (Rainio & Niemela 2003) or wētā in New Zealand (Watts et al. 2017). One of the reasons for this is because collecting and sorting invertebrates using conventional monitoring techniques is often expensive, time-consuming, and restricted by expertise in diagnostics. This limits our ability to assess the impacts of restoration manipulations such as pest control or habitat planting on invertebrate communities.

Emerging environmental DNA metabarcoding techniques could potentially revolutionise biodiversity monitoring for cryptic groups such as invertebrates by providing the ability to characterise entire communities from a single easily-collected environmental sample (Taberlet et al. 2012; Bohmann et al. 2014; Thomsen & Willerslev 2015; Holdaway et al. 2017). These techniques have the potential to make it easier and more cost-effective for conservation groups to undertake invertebrate community monitoring and to evaluate the performance of restoration projects. DNA metabarcoding analyses work by extracting DNA from environmental samples such as soil, water, leaves, or bulk invertebrates, and then amplifying and sequencing specific gene regions that can act as genetic barcodes to identify species whose DNA was present in the sample (e.g. Bohmann et al. 2014; Drummond et al. 2015; Holdaway et al. 2017; Yu et al. 2012). Soil is a well-known repository for DNA molecules from both soil-dwelling invertebrates (mites, nematodes, earthworms, larvae) and other terrestrial invertebrates (as dead carcases, exoskeletons, frass) (Drummond et al. 2015). As such, soil samples could provide an easily collected substrate for DNA-based analyses of invertebrate biodiversity. DNA extracted directly from soil will include a mixture of DNA molecules from living and dead organisms that reside in the soil (Carini 2016), and molecules from material deposited in the soil from elsewhere (e.g. the forest canopy). Furthermore, DNA metabarcoding analysis of bulk invertebrate samples (such as malaise trap collections) offers great potential to circumvent the need for time-consuming, costly, and demanding microscope-based identification of invertebrate specimens (Morinière et al. 2016; Yu et al. 2012).

DNA metabarcoding approaches could therefore be important monitoring tools for invertebrates within many restoration projects, particularly those at the landscape scale. However, there is a need for comparative analyses that apply both DNA metabarcoding and conventional

methods to the same taxonomic groups to help test and improve the accuracy and reliability of DNA-based methods and to assess the additional value obtained by utilising a DNA-based approach to invertebrate monitoring (Holdaway et al. 2017; Ji et al. 2013).

An example of a landscape-scale restoration project recently launched in New Zealand is the *Cape-to-City* project. It aims to restore indigenous biodiversity across 26,000 ha of productive landscape in Hawke's Bay, New Zealand, through the integrated use of wide-scale mammal predator control and habitat restoration (Norbury & McLennan 2015). Having the ability to monitor invertebrate communities using cost-effective techniques is paramount in such projects. Using this context for our case study, we aimed to characterise the invertebrate fauna of an isolated, coastal forest fragment within the *Cape-to-City* project boundary, using both conventional invertebrate monitoring and DNA metabarcoding approaches. The budget for each approach was identical (NZ$20,000 in 2015), so that we could directly compare and assess the data generated in terms of value for money. Two different sample types were analysed using DNA metabarcoding methods: 1) DNA extracted from bulk invertebrate samples collected using the malaise and pitfall traps at the same location; and 2) DNA extracted from soil. This design enabled comparison of soil and above-ground (pitfall and malaise trap) communities, as well as comparisons between conventional monitoring data and DNA metabarcoding data. Our sampling focused on identifying differences in invertebrate communities on the forest edge versus the forest interior, due to the well-known effects of fragmentation on many invertebrate groups (Didham 1997; Didham et al. 1998; Harris & Burns 2000). Our study also focused on beetles (Insecta: Coleoptera) because they are known to be sensitive to fragmentation (Ewers & Didham 2008), are involved in a range of ecological processes, their taxonomy is better known than other invertebrate groups, and they are easily sampled and identified using conventional approaches. Specifically, we hypothesised that (1) for a similar cost, metabarcoding analysis of DNA extracted from bulk invertebrate samples would provide greater taxonomic coverage (breadth of taxa and resolution within taxa) compared with conventional morphological identification; (2) metabarcoding analysis of bulk invertebrate DNA would be able to reproduce patterns of diversity and composition detected by conventional methods, and ecological variation in these patterns between forest interior and forest edge; and (3) metabarcoding analysis of DNA extracted from soil would show similar community patterns to the conventional trap samples, therefore indicating that soil is a reservoir of DNA from the broader invertebrate community, potentially negating the need for alternative sampling strategies.

Methods

Study area and sampling design

The study was carried out at Mohi Bush Scenic Reserve (61 ha; 39.85643S, 1769011E), a remnant of native forest in the Hawke's Bay, New Zealand. The fragment is approximately 1.5 km long by 325 m wide (Fig. S1) with the forest canopy being dominated by tawa (*Beilschmiedia tawa*), with scattered emergent podocarps (miro (*Prumnopitys ferruginea*) and matai (*Prumnopitys taxifolia*)) overtopping a diverse sub-canopy layer including the trees *Hedycarya arborea* and *Melicytus ramiflorus*.

A total of twelve 20 x 20 m plots were located at Mohi Bush Scenic Reserve: six along the forest-pasture edge ('edge' plots) and six within the interior (>100 m from the forest-pasture edge) of the forest fragment ('interior' plots; Fig. S1). Edge plots were positioned so that half the plot was in pasture and the remaining half was in the forest. For further details on study area and design see Appendix S1, Supporting Information.

Conventional invertebrate collection and identification

Invertebrates were sampled using a modified malaise trap (Watts et al. 2012, 2015) and pitfall traps and placed in the centre of each 20 x 20 m plot (Fig. S2). One malaise trap was placed in the centre of each plot (Fig. S2). At the edge plots the malaise trap was located in the forest (<3 m from the pasture-forest edge) so that it was not visible to the public from the pasture. The collecting jar of the malaise trap contained 150 ml of 50% monopropylene glycol and was orientated northward. Four pitfall traps (each a 105-mm diameter cup containing 75 ml of 50% monopropylene glycol), were dug into the ground 5 m away from the four corners of the malaise trap within each plot (Fig. S2). Traps were set for 1 month from 10 December 2015 to 7 January 2016. Invertebrates were preserved in 70% ethanol.

Captured invertebrates were sorted and counted to Order level using a binocular microscope. Beetles were selected for further identification and sorted on the basis of external morphology to recognised taxonomic units (hereafter referred to as species) and, where possible, given generic and species-level identifications by an expert diagnostician. After the invertebrates were sorted and counted, the specimens from each plot were pooled into a single bulk sample per plot for DNA analyses.

Soil sample collection

Soil samples were collected from the same twelve 20 x 20 m plots as used in the conventional invertebrate monitoring. A total of 24 soil cores (organic and mineral horizons; 4–6 kg in total) were collected from each plot with samples randomly located following standard sampling protocols (Fig. S2) using sterilised trowels. The soil samples were stored at 4 degrees and transferred to the lab within 5 days of collection for DNA extraction.

DNA analyses of bulk invertebrates and soil samples

Each bulk invertebrate sample was ground into a paste and homogenised, after which DNA was extracted from approx. 300 mg of the homogenised material using the Machery-Nagel NucleoSpin 96 Tissue extraction kit. Further details of the invertebrate homogenisation and DNA extraction protocols are described in Appendix S1.

The soil samples from each plot were homogenised, and a 10-g sample was taken for DNA extraction. DNA was extracted from each soil sample using PowerMax soil DNA extraction kits (MoBio), following standard protocol.

Invertebrate DNA from both the soil and bulk invertebrate samples was isolated and amplified by PCR using invertebrate-targeted primers mlCOIintF (Leray et al. 2013) and HCO2198 (Folmer et al. 1994), with sample-specific identification tags. The initial template amplification stage was carried out following the Touchdown protocol (Leray et al. 2013). This was then followed by the second stage of the MoTasp protocol (Clark et al. 2014) to add the barcodes and sequencing adaptors. The amplicons were sequenced by New Zealand Genomics Limited in a 2 x 250 bp Illumina MiSeq run.

Processing of DNA sequencing results

Raw sequence data were screened for sequencing adaptors and demultiplexed by sample, resulting in paired R1 and R2 fastq files for each sample. The R1 and R2 sequences were merged together and relabelled by sample, followed by trimming of primers, quality filtering, dereplication, clustering into OTUs at a 97% identity threshold, and mapping of sequences to OTUs, using cutadapt v1.11 (Martin 2011), USEARCH v9 (Edgar 2013), and VSEARCH (Rognes et al. 2016). Each OTU was assigned a taxonomic identity based on BLAST comparisons against a database consisting of metazoan COI sequences extracted from Genbank. Only OTUs with a best BLAST match of at least 80% over at least 250 bp to a sequence from an expected terrestrial invertebrate taxon were included in subsequent biodiversity analyses. Details of bioinformatic processing steps are provided in Appendix S1.

Data analysis

We calculated the number and taxonomic composition of sequences and OTUs detected by metabarcoding analysis of bulk invertebrate DNA, and the resolution of the taxonomic matches (i.e. how many OTUs could be identified to species, genus, family, and order). We compared this with the taxonomic composition and resolution of the same invertebrate samples determined using conventional methods. We also compared the taxonomic composition of soil DNA samples with the invertebrate sample composition.

The mean numbers of specimens, sequences or OTUs detected in different taxonomic orders were compared between forest edge and interior samples using two-sided Student's t-tests, with p-values adjusted for multiple comparisons using the false discovery rate method (Benjamini & Hochberg 1995). To compare the community composition and diversity patterns between forest edge and interior samples using conventional data, bulk invertebrate DNA, and soil DNA methods, we used non-metric multidimensional scaling ordinations to visualise sample similarity patterns, and Mantel and Procrustes tests to compare the observed patterns between datasets. This was done both for the overall communities (based on invertebrate specimen counts, bulk invertebrate DNA OTU abundances, and soil DNA invertebrate OTU abundances) and for beetle species and beetle OTUs detected within each DNA metabarcoding dataset. Ordinations were also generated for different taxonomic classes and orders within the bulk invertebrate DNA and soil DNA datasets. All data analyses were carried out using R 3.3.1 (R Core Team 2016), including the packages vegan (Oksanen et al. 2016) and ggplot2 (Wickham 2009).

Results

Conventional assessment of invertebrate community composition

A total of 7,503 invertebrate specimens were collected using conventional sampling methods (Table S1). Of these, 4,069 were collected at the forest edge and 3,434 were collected in the forest interior. Over 99% of the specimens were arthropods, including five Classes and at least 17 Orders, including 12 Orders of Insecta (Table S2; Fig. 1). The most abundant Orders were Diptera (1543), Collembola (1264), Lepidoptera (1204) and Coleoptera (935). The abundance of certain Orders differed between the edge and interior samples (Fig. 2). The mean abundances of Opiliones, Dipteran, and Orthopteran specimens in the edge samples were 5-fold, 1.6-fold, and 2.8-fold higher than in the interior samples according to two-sided Student's t-tests, respectively (t = 4.41, 5.56, and 4.11; P_{adj} = 0.036, 0.007, and 0.036; df = 10). Conversely, the mean abundance of Collembola specimens was 2.5-fold higher in interior samples than in edge samples (t = -10.31; $P_{adj} < 0.001$; df = 10).

A total of 34 beetle families, including 121 species were caught, of which Carabidae was most abundant (345) followed by Curculionidae (122), Staphylinidae (100), and Chrysomelidae (79) (Table S3, Fig. 3). Seventeen beetle families were represented by fewer than five specimens. No beetle families had significantly differing mean abundance or species richness in edge compared with interior samples.

Overall DNA sequencing and BLAST identification

A total of 3.9 million raw sequence reads were obtained from the bulk invertebrate DNA extracts, and 4.2 million from the soil DNA extracts, which together resulted in a total of 4,999 OTUs. Of these, 3,814 OTUs had a top BLAST match to a sequence from an expected terrestrial invertebrate organism, with match identities averaging 82.7% over 296 bp (range 67.8%–100%; Fig. S3). The top matches for the other 1,217 OTUs were sequences from non-terrestrial or non-invertebrate groups, including Cnidaria, Porifera, Echinodermata and Chordata, with a mean identity of 79.5% over 289 bp, and all matches ≥ 100 b.p. with ≤ 88.5% identity. Almost all of these OTUs (1,167) occurred only in the soil DNA results.

There were 2,590 OTUs with top matches to expected terrestrial invertebrate sequences of ≥ 80% identity over ≥ 250 bp, which were included in subsequent biodiversity analyses. Of these, 583 OTUs occurred only in the bulk invertebrate DNA results, 1,798 occurred only in the soil DNA results, and 209 occurred in both. Just 58 OTUs (2.2%) matched a pre-existing sequence from a terrestrial invertebrate known to occur in New Zealand with ≥ 97% identity over ≥ 250 bp (Table S4). These included four Annelida OTUs, four Collembola OTUs, seven Arachnida OTUs, a single OTU from each of Diplopoda, Gastropoda and Chromadorea, and a total of 47 genera.

Composition of bulk invertebrate DNA results

The metabarcoding analysis of DNA extracted from the bulk invertebrate specimens resulted in the identification of a broadly similar range of organisms as those identified by conventional methods (Table S2, Fig. 1). Five groups that were present among the conventionally identified specimens (all with low abundances) were absent from the bulk invertebrate DNA results (Pseudoscorpiones, Archaeognatha, Mantodea, Phasmatodea, and Platyhelminthes). Conversely, low numbers of sequences and OTUs were detected from Haplotaxida, Psocoptera, Thysanoptera, Trichoptera, Onychophora, and Nematoda, but these groups were absent from the conventional identification results. In addition, mites and myriapods were identified to Order-level based on the bulk invertebrate DNA results, compared with Class or Subphylum respectively based on the conventional identifications. The most abundant (sequence reads) were also the most species-rich (OTUs) groups. These were Diptera, Orthoptera, Lepidoptera, and Coleoptera. The differences in specimen abundances observed between edge and interior samples in conventional identifications were not clearly reflected in the bulk invertebrate DNA sequence abundances (Fig. 2). Orthoptera was the only Order for which mean sequence or OTU counts differed significantly between edge and interior samples, with higher sequences and OTUs in interior samples than edge samples (t = –3.70 and –4.66; p_{adj} = 0.004 and 0.002; *df = 10*), in contrast to the conventional data.

Fewer beetle families (18) were detected among the bulk invertebrate DNA results than the conventional beetle specimens (34), although the number of OTUs (107, with a total of 145,826 sequences) was similar to the number of species (121) (Table S3). There were 19 low-abundance beetle families detected in the conventional results that were missing from the bulk invertebrate DNA results, whereas the opposite was true of four families. The most abundant beetle families in the bulk invertebrate DNA results were broadly consistent with those among the conventional beetle specimens, although the number of beetle OTUs in these families usually exceeded the numbers of species identified by conventional means. Carabidae were most abundant, followed by Staphylinidae, Elateridae, Chrysomelidae, and Curculionidae (Fig. 3). No beetle families had significantly differing numbers of sequences or OTUs between edge or interior samples (Fig. 3).

Composition of soil DNA results

The metabarcoding analysis of soil DNA resulted in the detection of a broader range of taxa than the conventional specimens or bulk invertebrate DNA results (Table S2, Fig. 1 and 2). In addition to most of the invertebrate groups detected in results of the other methods, the soil DNA results included sequences and OTUs from Lumbriculida, Megaloptera, Siphonaptera, Rotifera, Tardigrada, and additional orders of mites, myriapods and nematodes. Haplotaxida and Gastropoda sequences and OTUs were much more abundant in the soil DNA results than the bulk invertebrate DNA results, whereas sequences and OTUs identified as Isopoda were absent from the soil DNA results (although sequences and OTUs identified as Malacostraca were present). The most abundant groups in the soil DNA results were Araneae, Lepidoptera, Haplotaxida, Hemiptera, and Diptera. With the exception of Megaloptera (detected in low abundance, with somewhat higher mean sequences and OTUs in edge samples than in interior samples), there were no significant differences in the mean abundance of soil DNA sequences or OTUs per Order in the edge or interior samples.

Nineteen beetle families and 90 beetle OTUs were identified among the soil DNA results, although the number of beetle sequences was only 3,877. There were 22 and 10 beetle families present among the conventional and bulk invertebrate DNA results respectively that were not detected in the soil DNA results, whereas there were six beetle families in the soil DNA results that were not detected in either the conventional results or the bulk invertebrate DNA results (Table S3). There was less consistency between the most abundant beetle families in the soil DNA results compared with the bulk invertebrate DNA or the conventional specimen results, with Nitidulidae most abundant (976 sequences, but only three OTUs), followed by Elateridae , Corylophidae , Carabidae , Staphylinidae, and Chrysomelidae. Much like the bulk invertebrate DNA results, no

beetle families had significantly differing numbers of sequences or OTUs between edge or interior samples (Fig. 3).

Community structure similarities between datasets

Multivariate ordinations of sample similarity indicated that the composition of invertebrate communities detected in forest edge samples differed from communities detected in forest interior samples, according to all sampling methods and analyses (Fig. 4). Similar trends were detected in the beetle community composition according to both conventional methods and bulk invertebrate DNA analysis, but this was not the case for beetles detected in the soil DNA dataset.

There was a significant correlation between the beetle specimens and the beetle OTUs detected in the bulk invertebrate DNA dataset, and a near-significant correlation between the overall invertebrate specimens and the bulk invertebrate DNA dataset (Table S5). In contrast, there was no evidence of a correlation between the conventional invertebrate specimens or bulk invertebrate DNA results and the soil DNA results.

Ordinations of other Orders within the bulk invertebrate and soil DNA datasets tended to indicate differing community composition in the forest edge and interior sites, albeit with some overlap (e.g. Gastropoda, Rotifera, Lepidoptera, and Annelida OTUs detected in the soil DNA; Fig. 5). The clearest discrimination of interior and edge samples was observed for Araneae and Opiliones OTUs detected in the soil DNA dataset, followed by Hymenoptera OTUs detected in both DNA datasets, and Diptera OTUs detected in the bulk invertebrate DNA dataset (Fig. 5).

Discussion

Can DNA metabarcoding be used as a monitoring technique for invertebrate communities?

Our results support the idea that DNA metabarcoding can be used as a tool to characterize whole invertebrate communities from either bulk invertebrate samples collected using malaise or pitfall traps (Ji et al. 2013; Morinière et al. 2016; Yu et al. 2012), or from soil samples (Drummond et al. 2015). However, the sampling method has very strong effects on the resulting community. DNA metabarcoding data provided information on a diverse array of invertebrate taxa – more than conventional analysis for the same cost, especially for the soil DNA analysis. In addition, the observed ecological patterns (i.e. separation of communities between forest edge and forest interior; Ewers & Didham 2008) in the bulk invertebrate DNA results were comparable to those obtained through conventional approaches, but the DNA data had a broader coverage and allowed for investigation of these edge effects for a diverse range of invertebrate taxa, not just beetles. The use of DNA metabarcoding methods to analyse bulk invertebrate samples (e.g. those collected using

malaise or pitfall traps) therefore has considerable potential as a tool for invertebrate biodiversity analysis (Ji et al. 2013; Morinière et al. 2016; Yu et al. 2012).

Orders observed in the bulk invertebrate DNA results but not the conventional analysis results likely represent fragments of material (or DNA molecules) that were unobserved, misidentifications, or unidentifiable by visual means. This may include organisms that occur on or inside other invertebrates, such as parasites, and gut contents. Conversely, invertebrate taxa that were not detected by the bulk invertebrate DNA analysis may be due to PCR primer mismatches, reference database limitations, or incorrect conventional identifications.

The widest range of taxa was identified by the soil DNA analysis, which is due to soil containing organisms and DNA molecules from a broader (and somewhat different) range of organisms than are sampled by above-ground pitfall and malaise traps. For example, the taxa detected by the soil DNA analysis included soil-dwelling groups such as mites, nematodes, rotifers, and tardigrades, which were largely absent from the conventional and bulk invertebrate DNA results. This broad sample of biodiversity detected by the soil DNA analysis is consistent with the diverse and complex composition soil communities (Bardgett 2002; Decaens et al. 2006; Giller 1996), suggesting that this approach provided an additional layer of ecological information beyond that provided by the collection of above-ground invertebrates. It is possible that soil invertebrate communities did not show a strong edge versus interior separation as the above-ground community did as they may be driven by different factors and have differing feedback mechanisms (Wardle et al. 2004 and references therein).

The broader ranges of taxa detected by the DNA metabarcoding analysis methods (especially the soil DNA analysis) is also partly due to the DNA-based identification of certain groups to lower taxonomic rank levels than was feasible using conventional methods. For example, mite and myriapod specimens were not identified beyond the ranks of Subclass (Acari) or Subphylum (Myriapoda), respectively, by the conventional analysis, whereas these groups were each represented by up to four different orders in the DNA metabarcoding analysis results. While these identifications are imperfect, they nonetheless demonstrate the ability of DNA-based methods to provide greater taxonomic depth and resolution than can be achieved by conventional identification methods.

What are the advantages and disadvantages of conventional sampling versus DNA metabarcoding methods for monitoring invertebrate communities?

The percentage costs of conventional invertebrate sampling were dominated by specimen sorting and morphological identifications, whereas the main costs of DNA metabarcoding methods were

molecular laboratory procedures (DNA extraction, PCR, and sequencing library preparation) and bioinformatic processing (Table S6). Soil has the advantage of being easy to sample in the field (collected during a one-off visit), whereas malaise or pitfall traps require specialized sampling technology and repeat visits to the sample location to establish and empty the traps. However, the quality and abundance of invertebrate DNA obtained from soil samples in our study was lower than for the bulk invertebrates, with many sequences and OTUs of uncertain identity, notably those most closely matching marine organisms (e.g. Hydrozoa). This may be due to sequence database limitations (Ekrem et al. 2007; Kvist 2013; Kwong et al. 2012). Further analyses with improved reference data are needed to clarify the identity of these OTUs.

Conventional invertebrate sampling methods and DNA metabarcoding approaches have complementary aspects, and combining the advantages of each technique will result in substantial enhancements to these research areas and to restoration ecology (Table S6). Conventional sampling provides reliable sampling of invertebrates grouped at high-level taxa, but genus and species delineations and identifications are difficult and time-consuming to establish, due to a scarcity of suitable expertise in diagnostics. Conventional sampling also provides specimens that can be verified and deposited into collections for future research (e.g. taxonomic revision of groups; Table S6). DNA metabarcoding, in contrast, allows for the rapid determination of OTUs (broadly equivalent to species) and taxonomic identification of OTUs (providing that suitable reference sequence data is available), and hence is well suited for efficient monitoring of invertebrate and soil biodiversity. Inadequate reference databases are currently a limitation for DNA-based identification of many invertebrate groups however (Ekrem et al. 2007; Kvist 2013; Kwong et al. 2012). In addition, PCR amplification biases make abundance/biomass inferences based on DNA metabarcoding data problematic (Elbrecht & Leese 2015; Pinto & Raskin 2012; Porazinska et al. 2009).

What are the next steps in developing DNA metabarcoding methods as a monitoring tool for invertebrates?

The accuracy and resolution of metabarcoding is currently limited by a lack of DNA reference data for New Zealand invertebrates. Indeed, very few of the OTUs detected in this case could be confidently identified beyond Order level, due to a paucity of invertebrate COI sequences available in reference sequence databases (Ekrem et al. 2007; Kvist 2013; Kwong et al. 2012). On the other hand, that some OTUs could be readily identified to genus and species level demonstrates the advantages of DNA-based identification methods, if relevant reference sequence data is available. Clearly, further efforts to improve the coverage of sequence databases are needed if the full potential of DNA metabarcoding methods for invertebrate biodiversity monitoring is to be realized.

This study represents one of the first attempts to apply DNA metabarcoding methods to the monitoring of New Zealand invertebrate communities. Despite the current limitations of reference databases, and uncertainty about the impacts of PCR biases, these methods can efficiently provide immense data on a very wide range of biota, complementing conventional invertebrate community monitoring. It provided data on invertebrate groups (e.g. earthworms in the soil DNA) that are neglected but have significant roles in restoring ecosystems (Snyder and Hendrix 2008). Soil communities represent one of the greatest pools of unknown invertebrate diversity, and we suggest future DNA barcoding (and taxonomic discovery) efforts should be focused on soil biota.

The information gained during our study presented the opportunity to apply both DNA metabarcoding and conventional methodologies to the same taxonomic groups at the same site and in doing so we have clarified the accuracy and consistency of DNA-based techniques. DNA metabarcoding methods offer restoration managers who are frequently resource-limited a practical, cost-effective technique for monitoring invertebrate communities. Future research could focus on using this technique in an applied restoration setting to determine the success of restoring invertebrate communities.

Acknowledgements

This research was supported by SSIF funding to Landcare Research from the Ministry of Business, Innovation and Employment's Science and Innovation group and the NZ's Biological Heritage National Science Challenge. In addition, Hawke's Bay Regional Council and the Aotearoa Foundation contributed funding to the research as part of the Cape-to-City project. Department of Conservation (Napier) gave permission for this survey to occur and we thank Dave Carlton in particular for assistance with iwi consultation. We thank Robbie Price for help with selecting plot location and drawing Figure S1, and Karen Boot for helping to process the DNA samples. Thanks to Stephen Thorpe for identifying the beetles. Thomas Buckley and Anne Austin provided useful comments on the draft manuscript.

LITERATURE CITED

Bardgett RD (2002) Causes and consequences of biological diversity in soil. Zoology 105(4):367-374

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological) 57: 289-300

Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, Knapp, M, Yu DW, de Bruyn M (2014) Environmental DNA for wildlife biology and biodiversity monitoring. Trends In Ecology & Evolution 29: 358-367

Carini P, Marsden PJ, Leff JW, Morgan EE, Strickland MS, Fierer N (2016) Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. Nature Microbiology 19: 16242

Clarke LJ, Czechowski P, Soubrier J, Stevens MI, Cooper A (2014) Modular tagging of amplicons using a single PCR for high-throughput sequencing. Molecular Ecology Resources 14: 117-121

Decaëns T, Jiménez JJ, Gioiac C, Measey GJ, Lavelle P (2006) The values of soil animals for conservation biology. European Journal of Soil Biology 42: S23-S38

Didham RK (1997) An overview of invertebrate responses to habitat fragmentation. Pages 303–320 In: Watt AD, Stork NE, Hunter MD (eds) Forests and insects. Chapman and Hall, London

Didham RK, Hammond PM, Lawton JH, Eggleton P, Stork NE (1998) Beetle species responses to tropical forest fragmentation. Ecological Monographs 68: 295-323

Drummond AJ, Newcomb RD, Buckley TR, Xie D, Dopheide A, Potter BCM, Heled J, Ross HA, Tooman L, Grosser S, Park D, Demetras NJ, Stevens MI, Russell JC, Anderson SH, Carter A, Nelson N (2015) Evaluating a multigene environmental DNA approach for biodiversity assessment. GigaScience 4: 46

Edgar R (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature Methods 10: 996-998

Ekrem T, Willassen E, Stur E (2007) A comprehensive DNA sequence library is essential for identification with DNA barcodes. Molecular Phylogenetic and Evolution 43: 530-542

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* 10.7: e0130324

Ewers RM, Didham RK (2008) Pervasive impact of large-scale edge effects on a beetle community. Proceedings of the National Academy of Sciences 105: 5426-5429

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3: 294-299

Forup ML, Memmott J (2005) The restoration of plant-pollinator interactions in hay meadows. Restoration Ecology 13: 265-274

Gerlach J, Samways M, Pryke J (2013) Terrestrial invertebrates as bioindicators: an overview of available taxonomic groups. Journal of Insect Conservation 17: 831-850

Giller P (1996) The diversity of soil communities, the 'poor man's tropical rainforest'. Biodiversity and Conservation 5: 135-168

Gregoire Taillefer A, Wheeler TA (2012) Community assembly of Diptera following restoration of mined boreal bogs: taxonomic and functional diversity. Journal of insect Conservation 16: 165-176

Harris RJ, Burns BR (2000) Beetle assemblages of kahikatea forest fragments in a pasture-dominated landscape. New Zealand Journal of Ecology 24: 57-68

Holdaway R, Wood, J, Dickie I, Orwin K, Bellingham P, Richardson S, Lyver P, Timoti P, Buckley TR (2017) Using DNA metabarcoding to assess new Zealand's terrestrial biodiversity. New Zealand Journal of Ecology 41: DOI: 10.20417/nzjecol.41.28

Ji Y, Ashton L, Pedley SM, Edwards DP, Tang Y, Nakamura A, Kitching R, Dolman PM, Woodcock P, Edwards FA, Larsen TH, Hsu WW, Benedick S, Hamer KC, Wilcove DS, Bruce C, Wang X, Levi T, Lott M, Emerson BC, Yu DW (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. Ecology Letters 16: 1245-1257

Kvist S (2013) Barcoding in the dark? A critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge. Molecular Phylogenetics and Evolution 69: 39-45

Kwong S, Srivathsan A, Meier R (2012) An update on DNA barcoding: Low species coverage and numerous unidentified sequences. Cladistics 28: 639-644

Lavelle P, Decaëns T, Aubert M, Barot S, Blouin M, Bureau F, Margerie P, Mora P, Rossi JP (2006) Soil invertebrates and ecosystem services. European Journal of Soil Biology 42: S3-15

Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Frontiers in Zoology 10: 1-14

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10-12. DOI: http://dx.doi.org/10.14806/ej.17.1.200

Morinière J, Cancian de Araujo B, Lam AW, Hausmann A, Balke M, Schmidt S, Hendrich L, Doczkal D, Fartmann B, Arvidsson S, Haszprunar G (2016) Species identification in malaise trap samples by DNA barcoding based on NGS technologies and a scoring matrix. Plos One 11: e0155497.

Norbury G, McLennan J 2015. Biodiversity and predator monitoring for *Cape-to-City*, Hawke's Bay Project. Landcare Research Contract Report LC2237

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2016). vegan: Community Ecology Package. R package version 2.4-1. https://CRAN.R-project.org/package=vegan

Pinto A, Raskin L (2012) PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. PLoS ONE: e43093

Porazinska DL, Giblin-Davis RM, Faller L, Farmerie W, Kanzaki N, Morris K, Powers T, Tucker AE, Sung W, Thomas WK (2009) Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. Molecular Ecology Resources 9(6): 1439-1450

R Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rainio J, Niemela J (2003) Ground beetles (Coleoptera: Carabidae) as bioindicators. Biodiversity and Conservation 12: 487-506

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ 4: e2584.

Snyder BA, Hendrix PF (2008) Current and potential roles of soil macroinvertebrates (earthworms, millipedes, and isopods) in ecological restoration. Restoration Ecology 16: 629-36

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21: 2045-2050

Thomsen PF, Willerslev E (2015) Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. Table

Biological Conservation 183: 4-18

Wardle DA, Bardgett RD, Klironomos JN, Setälä H, Van Der Putten WH, Wall DH (2004) Ecological linkages between aboveground and belowground biota. Science 304(5677): 1629-33

Watts C, Clarkson BR, Didham RK (2008) Rapid beetle community convergence following experimental habitat restoration in a mined peat bog. Biological Conservation 141: 568-579

Watts C, Stringer I, Innes J, Monks J (2017) Evaluating tree wētā (Orthoptera: Anostostomatidae: *Hemideina* species) as bioindicators for New Zealand national biodiversity monitoring. Journal of Insect Conservation DOI 10.1007/s10841-017-9997-8

Wickham H (2009) ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York

Yu DW, Yinqiu J, Emerson BC, Xiaoyang W, Chengxi Y, Chunyan D, Zhaoli D (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods in Ecology and Evolution 3: 613-623

Figure captions

Figure 1 Overall taxonomic composition of invertebrate specimens, and DNA sequence reads derived from DNA metabarcoding analysis of bulk invertebrate DNA and soil DNA extracts. Taxonomic groups are Orders, except where indicated by parentheses.

Figure 2 Mean numbers of invertebrate specimens detected per Order by conventional methods, and invertebrate OTUs detected per Order by DNA metabarcoding methods, in forest edge and interior samples. Only orders represented by at least five mean specimens or OTUs per sample are included. Error bars represent 95 % confidence intervals.

Figure 3 Mean numbers of beetle species and OTUs per Family detected in forest edge and interior samples using conventional invertebrate sampling and DNA metabarcoding methods. Only families represented by at least three mean species or OTUs per sample are included. Error bars represent 95 % confidence intervals.

Figure 4 Non-metric MDS ordinations based on invertebrate specimens, bulk invertebrate DNA and soil DNA metabarcoding results from forest edge and interior samples, and the beetle components of these datasets.

Figure 5 Non-metric MDS ordinations based on various components of the bulk invertebrate DNA and soil DNA metabarcoding datasets from forest edge and interior samples.

Figure 1

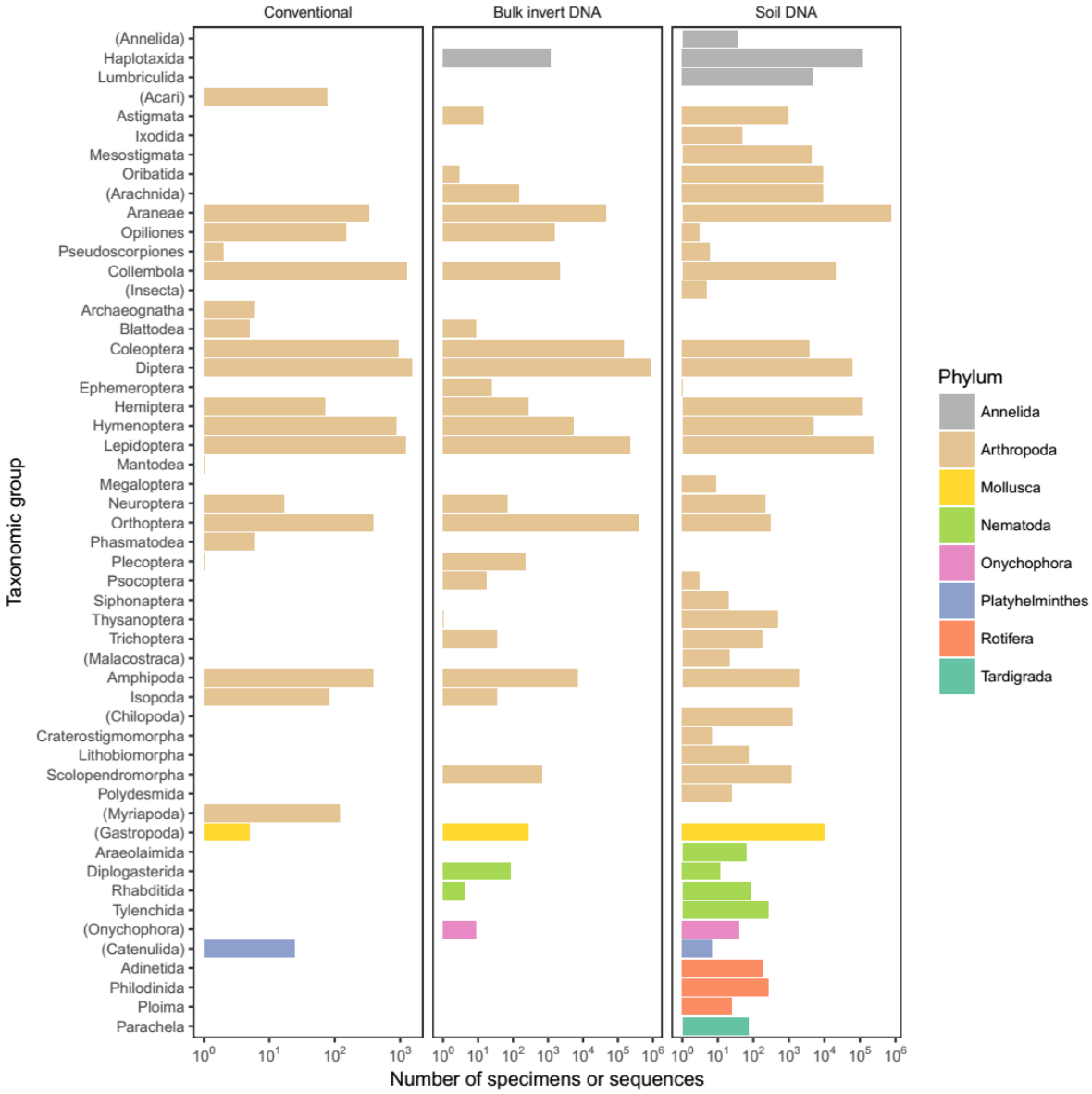

Conventional
Bulk invert DNA
Soil DNA
Taxonomic group
(Annelida)
Haplotaxida
Lumbriculida
(Acari)
Astigmata
Ixodida
Mesostigmata
Oribatida
(Arachnida)
Araneae
Opiliones
Pseudoscorpiones
Collembola
(Insecta)
Archaeognatha
Blattodea
Coleoptera
Diptera
Ephemeroptera
Hemiptera
Hymenoptera
Lepidoptera
Mantodea
Megaloptera
Neuroptera
Orthoptera
Phasmatodea
Plecoptera
Psocoptera
Siphonaptera
Thysanoptera
Trichoptera
(Malacostraca)
Amphipoda
Isopoda
(Chilopoda)
Craterostigmomorpha
Lithobiomorpha
Scolopendromorpha
Polydesmida
(Myriapoda)
(Gastropoda)
Araeolaimida
Diplogasterida
Rhabditida
Tylenchida
(Onychophora)
(Catenulida)
Adinetida
Philodinida
Ploima
Parachela
10^0 10^1 10^2 10^3 10^4 10^5 10^6
Number of specimens or sequences
Phylum
Annelida
Arthropoda
Mollusca
Nematoda
Onychophora
Platyhelminthes
Rotifera
Tardigrada

Figure 2

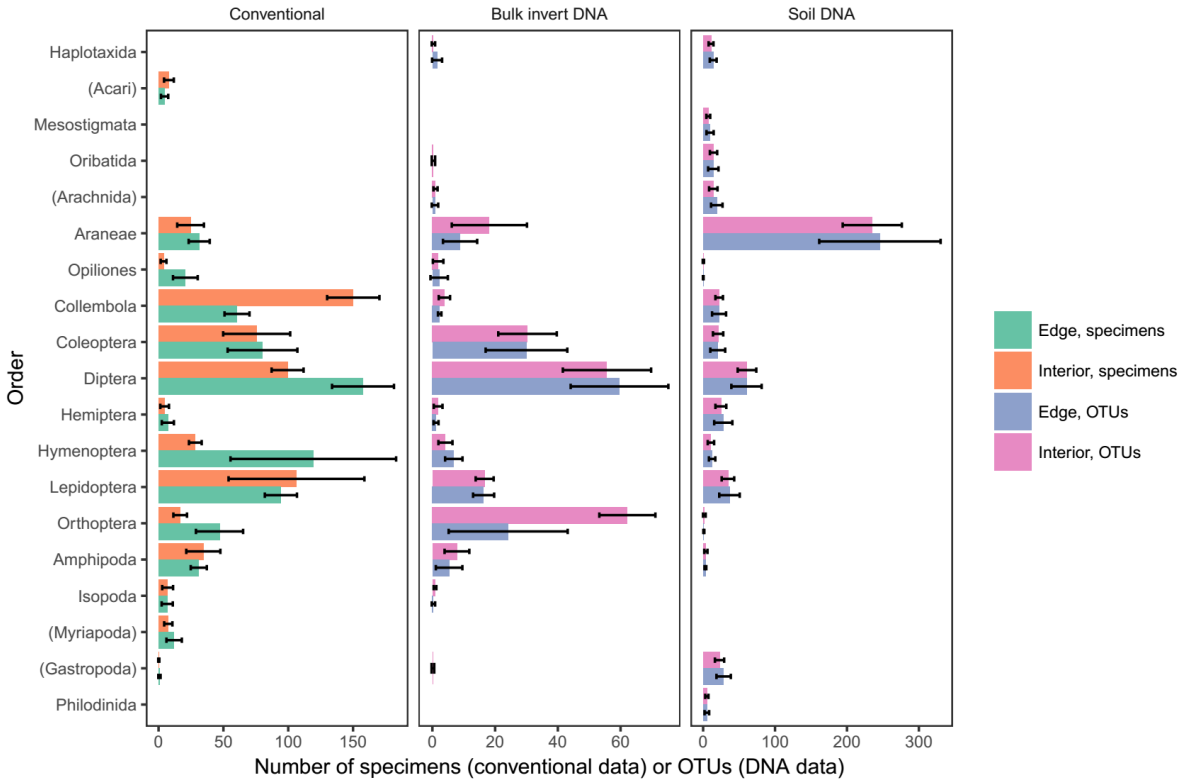

Conventional
Bulk invert DNA
Soil DNA
Order
Haplotaxida
(Acari)
Mesostigmata
Oribatida
(Arachnida)
Araneae
Opiliones
Collembola
Coleoptera
Diptera
Hemiptera
Hymenoptera
Lepidoptera
Orthoptera
Amphipoda
Isopoda
(Myriapoda)
(Gastropoda)
Philodinida
0
50
100
150
0
20
40
60
0
100
200
300
Number of specimens (conventional data) or OTUs (DNA data)
Edge, specimens
Interior, specimens
Edge, OTUs
Interior, OTUs

Figure 3

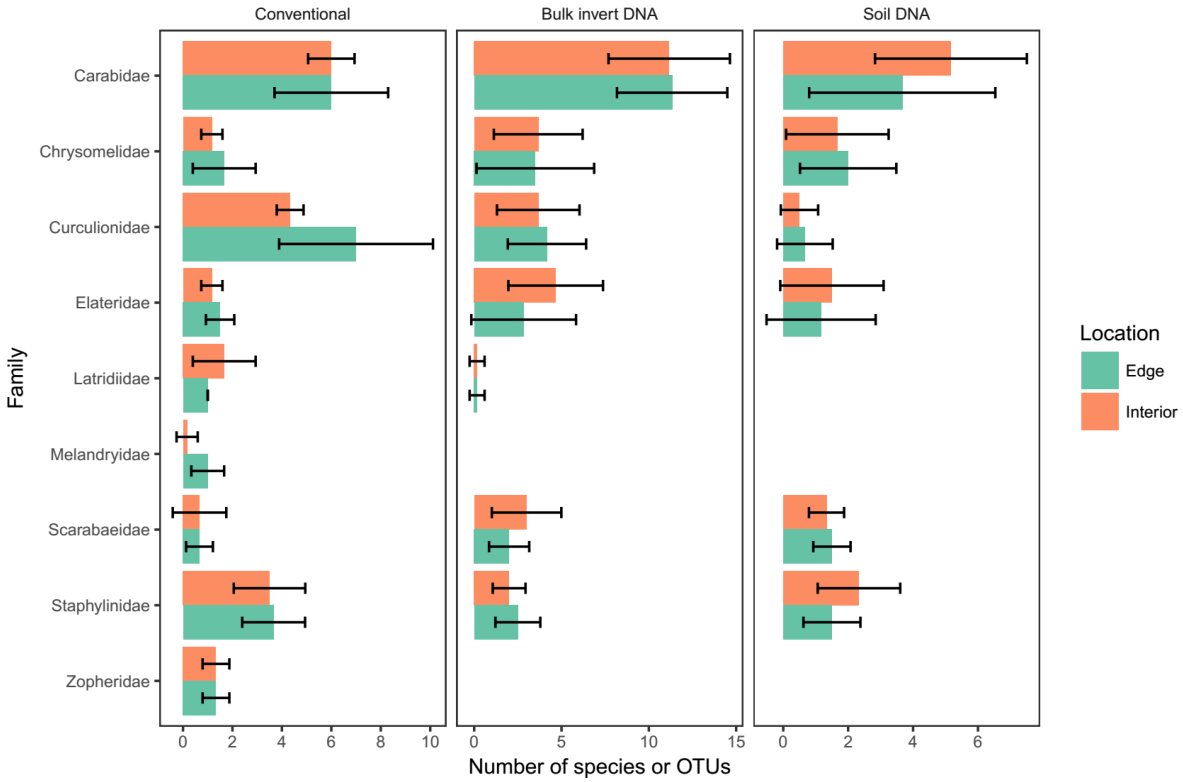

Conventional
Bulk invert DNA
Soil DNA
Carabidae
Chrysomelidae
Curculionidae
Elateridae
Latridiidae
Melandryidae
Scarabaeidae
Staphylinidae
Zopheridae
Family
0
2
4
6
8
10
0
5
10
15
0
2
4
6
Number of species or OTUs
Location
Edge
Interior

Figure 4

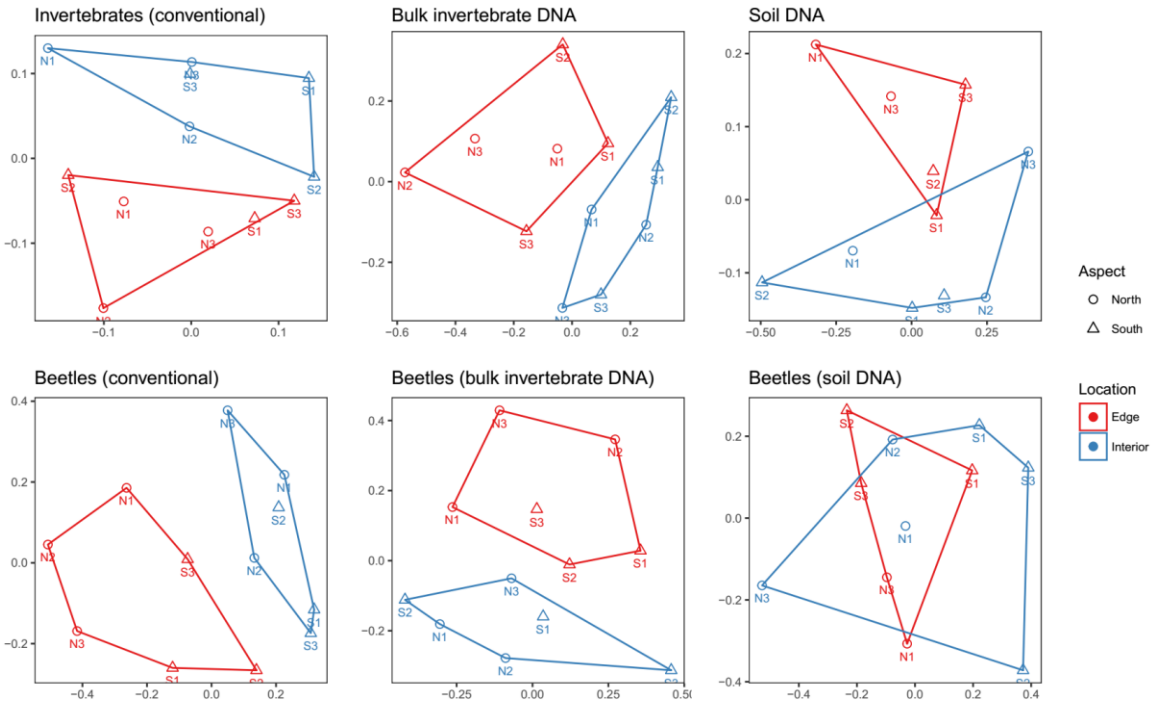
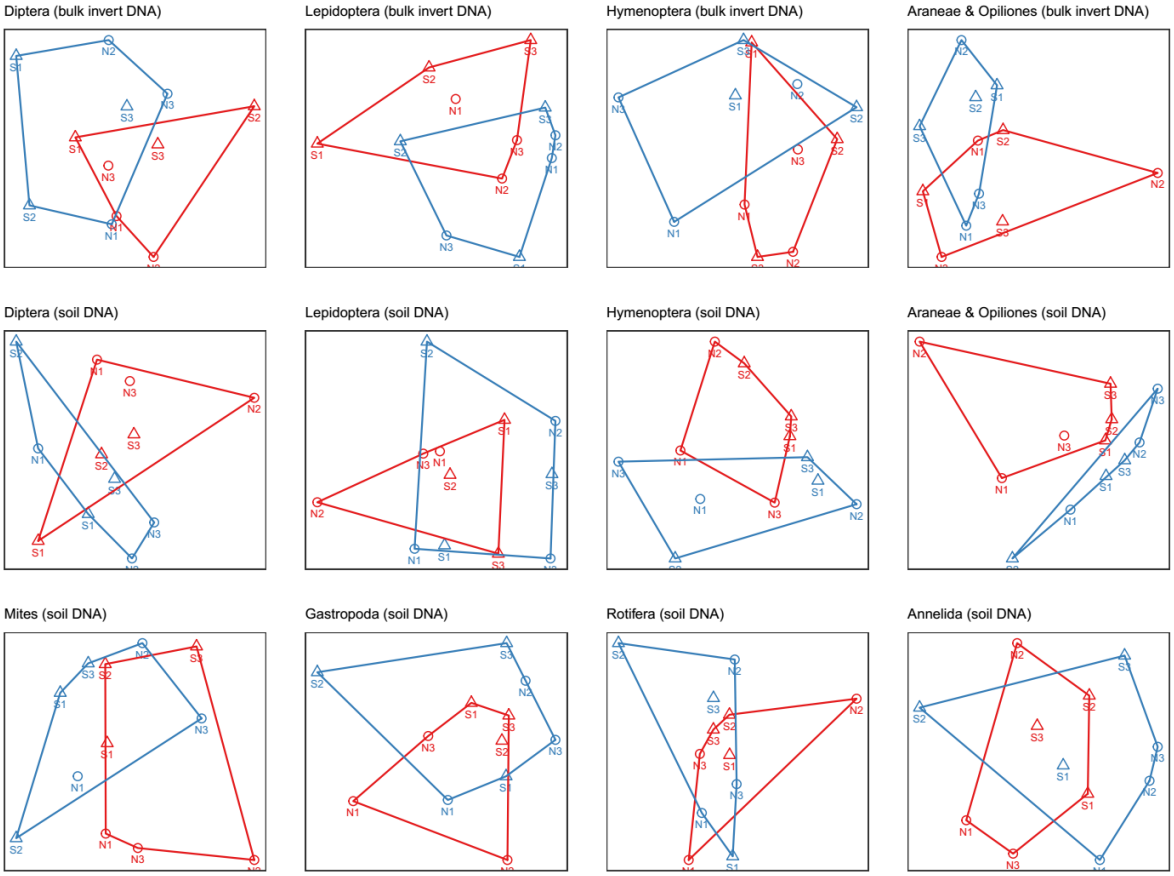

Invertebrates (conventional)
Bulk invertebrate DNA
Soil DNA
Beetles (conventional)
Beetles (bulk invertebrate DNA)
Beetles (soil DNA)
Aspect
North
South
Location
Edge
Interior

Figure 5


Diptera (bulk invert DNA)
Lepidoptera (bulk invert DNA)
Hymenoptera (bulk invert DNA)
Araneae & Opiliones (bulk invert DNA)
Diptera (soil DNA)
Lepidoptera (soil DNA)
Hymenoptera (soil DNA)
Araneae & Opiliones (soil DNA)
Mites (soil DNA)
Gastropoda (soil DNA)
Rotifera (soil DNA)
Annelida (soil DNA)

Appendix S1. Supporting Information.

Additional details regarding study area and design

Mohi Bush Scenic Reserve is one of the surviving native forest fragments surrounded within the productive landscapes of the Hawke's Bay region. The remnant has been previously partially logged with the canopy dominated by tawa and scattered podocarps. There was a diverse sub-canopy layer including the trees *Hedycarya arborea*, *Melicytus ramiflorus*, and Coprosma grandiflora, and the treefern *Dicksonia squarrosa*. Small native ferns, vines, liverworts, and mosses were also common. The fragment was surrounded by pasture dominated by a diverse array of introduced grasses, such as *Lolium perenne*, *Dactylis glomerata*, *Holcus lanatus*, and *Bromus diandrus*.

Potential edge plot locations were determined by placing points 25 m apart along the northern and southern pasture-forest boundary of Mohi Bush Scenic Reserve using an aerial photograph and a programme implemented through a purpose-built extension to ArcView 3.2 (ESRI 1999). In the field, edge plots were rejected if they 1) were less than 120 m apart, 2) contained forest trees that had fallen into the pasture interrupting the abrupt pasture-forest edge, or 3) contained large amounts of ongaonga (*Urtica ferox*, a highly toxic nettle). Three edge plots were located along the northern ('N' plots) and 3 along the southern ('S' plots) boundaries of Mohi Bush Scenic Reserve (Fig. S1). From each edge plot, the interior plots were situated by moving 100 m on a bearing into the forest. The bearing for the N plots was 192° and for the S plots was 9°. As each interior plot had to be at least 120 m apart, the bearings were different for the N and S plots. Due to two interior plots being positioned on and very near tracks, the bearing and distance into the forest were slightly adjusted (forest N3 bearing 203°, distance 130 m; forest S1 bearing 6°, distance 131 m; Fig. S1). Excepting for these adjustments, plot locations were entirely objectively determined. The location of each plot was recorded using a Garmin 60CSx GPS.

Figure S1. Map of Mohi Bush Scenic Reserve in Hawke's Bay, New Zealand showing the locations of sampling plots.

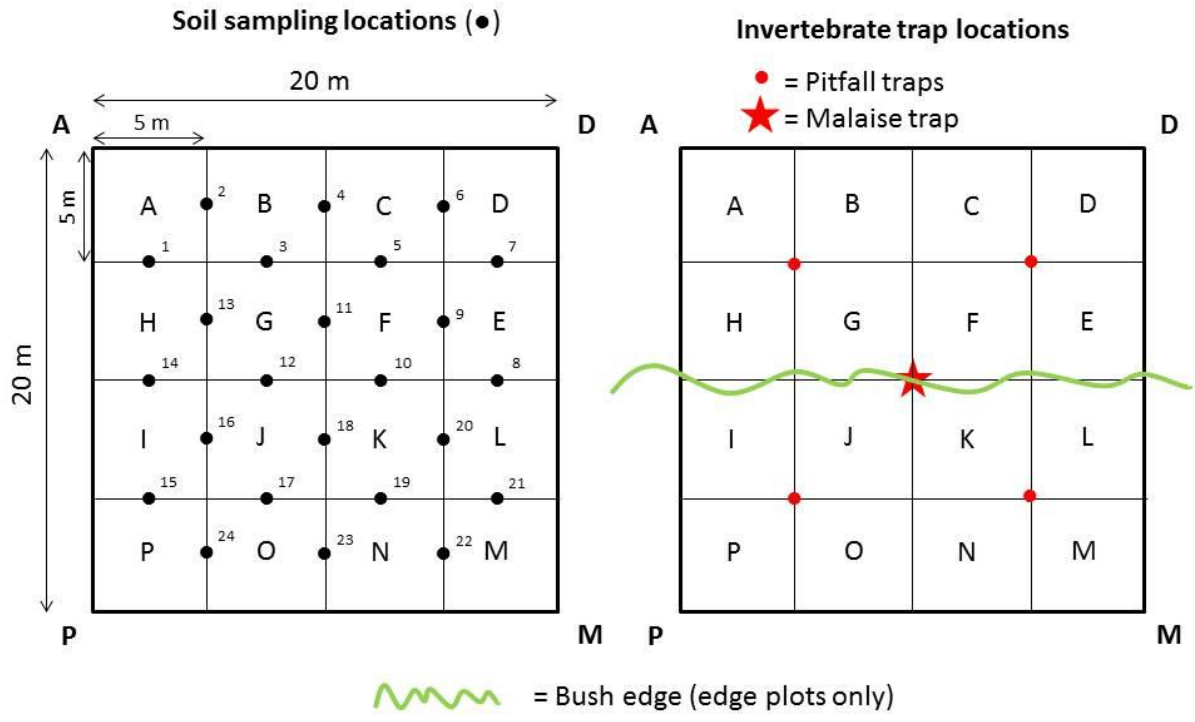


Figure S2. Layout of sampling points within the 20 x 20 m plots.

Additional details regarding DNA extraction protocols

DNA was extracted from a single bulk invertebrate sample per plot, consisting of the invertebrates collected from both the malaise traps and the pitfall traps located at each plot. This was the same sample that underwent conventional identification, and this allowed direct comparison of the conventional invertebrate monitoring results. The invertebrate samples were homogenised using the following procedure:

1. Remove EtOH from all sub-samples from the same plot, by evaporation and/or use of a Pasteur pipette.
2. Spray the bench with bleach solution (10%) and wipe clean.
3. Spray the bench with ethanol (70%) and wipe clean.
4. Spray a pestle and mortar with TriGene, wipe with a tissue to ensure all surfaces are covered, leave for 1-2 min, and then wipe away any remaining liquid.
5. Spray the pestle and mortar with bleach solution (10%), wipe with a tissue to ensure all surfaces are covered, leave for 1-2 min, and then wipe away any remaining liquid.
6. Spray the pestle and mortar with ethanol (70%) and allow to evaporate dry (complete drying at this stage is not essential).
7. Place several metal spatulas into a tube of bleach, and swirl around for 1-2 min.
8. Spray the spatulas with ethanol (70%) and allow to evaporate dry (complete drying at this stage is not essential).
9. Sterilise all the cleaned equipment with UV. Once complete, place the equipment on paper towels on the cleaned bench.
10. Fill a small container with liquid nitrogen. Pour a small amount into the mortar and allow both the pestle and mortar to cool.
11. Transfer all invertebrate sub-samples from the same plot into the mortar. Any remaining EtOH will hiss and pop a bit, so beware. Pour on more liquid nitrogen as required to freeze all the invertebrate material.
12. Bang and grind the invertebrate material until it is a fine powder. Add more liquid nitrogen as required.
13. Transfer the homogenized material into a 50-ml centrifuge tube, using sterilised metal spatulas.
14. Flash freeze the sample tube in liquid nitrogen, then transfer to a -80°C freezer.
15. Wipe the mortar and pestle clean with a tissue, followed by a rinse with water and detergent.
16. Repeat the process for all remaining samples.

DNA was extracted from approx. 300 mg of each homogenised sample using the Machery-Nagel NucleoSpin 96 Tissue extraction kit. The initial lysis step used 400 ul of digestion buffer, consisting of 10 mM Tris (pH 8), 10 mM NaCl, 5 mM $CaCl_2$, 2.5 mM EDTA (pH 8), 2% SDS, 40 mM DTT, and 2 mg/ml proteinase K, and was left to incubate overnight at 56^{o} with shaking. The remainder of the extraction

protocol followed the manufacturer's instructions, but with 500 ul of Binding Buffer BQ1 and 500ul of absolute EtOH added to each sample to adjust the DNA binding conditions (step 3). The automated parts of the protocol were carried out on the Janus Automated Liquid Handling System (PerkinElmer).

Additional details regarding processing of DNA sequencing results

Raw sequence data was screened for sequencing adaptors and demultiplexed by sample, resulting in paired R1 and R2 fastq files for each sample. The R1 and R2 sequences were merged together using -fastq_mergepairs in USEARCH v9 (Edgar 2013). Amplicon primers were trimmed from the merged sequences using cutadapt v1.11 (Martin 2011). The sequences were then quality filtered and dereplicated in VSEARCH (Rognes 2016) using -fastq_filter with maximum expected errors (-fastq_maxee) of 1.0, and -derep_fulllength with minimum sequence length (-minseqlength) of 200 bp., respectively. The filtered and dereplicated sequences were clustered into OTUs at a 97 % identity threshold with a minimum OTUabundance of 2 sequences (-minsize 2), using the UPARSE algorithm (-cluster_otus) within USEARCH v9 (Edgar 2013). Finally, the primer-trimmed and merged sequences were mapped to the OTU sequences at an identity threshold of 97 % using -usearch_global in VSEARCH (Rognes et al. 2016).

Each OTU was assigned a taxonomic identity based on BLAST comparisons against a database consisting of metazoan COI sequences. The database was compiled by downloading all sequences returned by a search of DDBJ using the following criteria:

Molecule type: DNA

Sequence length: 200-5000

Definition: (COI OR "CO1" OR "cytochrome oxidase" OR "COX1" OR "cytochrome c oxidase") NOT ("unverified" OR "uncultured")

Lineage: Metazoa

The downloaded sequences were filtered to contain only those with a valid genus name, resulting in 982,697 sequences. The top BLAST match for each OTU was used to infer its identity.

The OTUs were filtered based on the quality and identity of their BLAST matches. Only OTUs with a best BLAST match of at least 80 % over at least 250 bp to a sequence from an expected terrestrial invertebrate taxon were included in subsequent biodiversity analyses. OTUs were retained if the identity of the top BLAST match was any of the following partially or wholly terrestrial invertebrate taxa: Annelida, Arachnida, Collembola, Insecta, Malacostraca, Gastropoda, Myriapoda, Nematoda, Onychophora, Platyhelminthes, Protura, Rotifera ,and Tardigrada (see Fig. S3). These OTUs were further filtered and only retained if the identity of the top BLAST match was not any of the following

non-terrestrial Annelida, Malacostraca, or Gastropoda groups: Bivalvia, Cephalopoda, Decapoda, Mysida, Nudibranchia, or Polychaeta. Any sequence reads occurring in the negative control samples were subtracted from their corresponding OTUs in all other samples.

Table S1. The number of invertebrate specimens, OTUs, sequences and orders, and beetle specimens, species, OTUs, sequences and families detected in edge and interior samples according to conventional invertebrate analysis methods, bulk invertebrate DNA sequencing, and soil DNA sequencing

Analysis method	Component	Location	Total	Mean	SD	Invertebrate orders or beetle families
Conventional	All specimens	Edge	4,069	678	95.5	21
		Interior	3,434	572	58.2	21
	Beetle specimens	Edge	475	79.2	25.8	31
		Interior	447	74.5	25.2	25
	Beetle species	Edge	96	32.8	8.33	
		Interior	69	26.8	5.04	
Bulk invertebrate DNA	Invertebrate OTUs	Edge	523	171	55.4	22
		Interior	603	215	33.8	26
	Beetle OTUs	Edge	85	30.3	12.7	17
		Interior	81	30.5	9.03	12
	Invertebrate sequences	Edge	1,317,544	219,591	41,101	
		Interior	1,247,733	207,956	18,005	
	Beetle sequences	Edge	63,745	10,624	9,667	
		Interior	82,081	13,680	15,231	
Soil DNA	Invertebrate OTUs	Edge	1,505	542	177	36
		Interior	1,516	512	91.4	35
	Beetle OTUs	Edge	69	20.3	9.58	18
		Interior	69	20.5	6.66	16
	Invertebrate sequences	Edge	842,893	140,482	81,558	
		Interior	526,490	87,748	43,564	
	Beetle sequences	Edge	2,048	341	478	
		Interior	1,829	305	380	

Table S2. Order-level taxonomic composition and abundance of invertebrate specimens collected by conventional methods, and DNA sequence and OTUs from bulk invertebrates and soil DNA samples, from forest edge and interior sites

Phylum	Class	Order	Specimens, Edge	Specimens, Interior	Bulk invert DNA, Edge, OTUs	Bulk invert DNA, Edge, sequences	Bulk invert DNA, Interior, OTUs	Bulk invert DNA, Interior, sequences	Soil DNA, Edge, OTUs	Soil DNA, Edge, sequences	Soil DNA, Interior, OTUs	Soil DNA, Interior, sequences
Annelida	(Annelida)	(Annelida)							4	24	2	12
Annelida	(Annelida)	Haplotaxida			5	1152	2	4	38	75409	38	50460
Annelida	(Annelida)	Lumbriculida							1	314	2	4318
Arthropoda	Arachnida	(Acari)	28	48								
Arthropoda	Arachnida	(Arachnida)			4	136	3	18	53	7143	41	2108
Arthropoda	Arachnida	Araneae	188	148	38	45272	64	1729	641	538132	643	204137
Arthropoda	Arachnida	Astigmata					1	14	4	958		
Arthropoda	Arachnida	Ixodida							2	49		
Arthropoda	Arachnida	Mesostigmata							31	2492	24	1728
Arthropoda	Arachnida	Opiliones	124	23	13	3171	8	27843	2	2	3	3
Arthropoda	Arachnida	Oribatida					2	3	46	2488	46	6531
Arthropoda	Arachnida	Pseudoscorpiones	1	1							1	6
Arthropoda	Chilopoda	(Chilopoda)							1	1172	1	93
Arthropoda	Chilopoda	Lithobiomorpha							1	75		
Arthropoda	Chilopoda	Scolopendromorpha			3	684	1	3	4	1037	5	143
Arthropoda	Diplopoda	Julida			1	79	1	193	1	10	1	126
Arthropoda	Diplopoda	Polydesmida							1	2	3	23
Arthropoda	Entognatha	Collembola	363	901	6	1468	10	699	76	23642	72	10477
Arthropoda	Insecta	Archaeognatha	3	3								
Arthropoda	Insecta	Blattodea	2	3			2	9				
Arthropoda	Insecta	Coleoptera	481	454	85	63745	81	82081	69	2048	69	1829
Arthropoda	Insecta	Diptera	946	597	177	500266	171	510029	171	46451	183	14335
Arthropoda	Insecta	Ephemeroptera			1	5	1	19	1	1		
Arthropoda	Insecta	Hemiptera	43	28	7	49	9	237	84	102125	76	14345
Arthropoda	Insecta	Hymenoptera	716	170	24	4475	16	1088	31	4326	38	539
Arthropoda	Insecta	Lepidoptera	566	638	46	619427	49	300106	108	29784	112	202953
Arthropoda	Insecta	Mantodea	1									
Arthropoda	Insecta	Megaloptera							1	8	1	1
Arthropoda	Insecta	Neuroptera	9	8	1	4	1	65	3	92	5	126
Arthropoda	Insecta	Orthoptera	282	100	89	76190	158	316888	4	106	10	204
Arthropoda	Insecta	Phasmatodea	2	4								
Arthropoda	Insecta	Plecoptera		1	1	4	2	229				

Phylum	Class	Order	Specimens, Edge	Specimens, Interior	Bulk invert DNA, Edge, OTUs	Bulk invert DNA, Edge, sequences	Bulk invert DNA, Interior, OTUs	Bulk invert DNA, Interior, sequences	Soil DNA, Edge, OTUs	Soil DNA, Edge, sequences	Soil DNA, Interior, OTUs	Soil DNA, Interior, sequences
Arthropoda	Insecta	Psocoptera			1	17	2	116			2	2113
Arthropoda	Insecta	Siphonaptera							1	10	1	10
Arthropoda	Insecta	Thysanoptera			1	1			1	6	1	501
Arthropoda	Insecta	Trichoptera					1	36	5	97	2	75
Arthropoda	Malacostraca	(Malacostraca)							1	12	1	9
Arthropoda	Malacostraca	Amphipoda	186	207	15	1346	13	5963	9	767	12	1637
Arthropoda	Malacostraca	Isopoda	40	42	1	5	1	29				
Arthropoda	Myriapoda	(Myriapoda)	72	45								
Mollusca	Gastropoda	(Gastropoda)	4	1	1	5	1	276	67	3490	73	7248
Nematoda	Chromadorea	Araeolaimida							2	55	2	7
Nematoda	Chromadorea	Diplogasterida			1	41	1	45	1	2	1	10
Nematoda	Chromadorea	Rhabditida			2	2	1	2	6	59	4	23
Nematoda	Chromadorea	Tylenchida							3	241	2	19
Onychophora	(Onychophora)	(Onychophora)					1	9	3	8	3	32
Platyhelminthes	Catenulida	(Catenulida)	12	12							1	7
Rotifera	Bdelloidea	Adinetida							12	78	16	117
Rotifera	Bdelloidea	Philodinida							12	129	14	137
Rotifera	Monogononta	Ploima							2	9	3	16
Tardigrada	Eutardigrada	Parachela							2	40	2	32
Annelida	(Annelida)	(Annelida)							4	24	2	12

Table S3. Family-level taxonomic composition and abundance of Coleoptera specimens collected by conventional methods, and Coleoptera DNA sequences and OTUs from bulk invertebrates and soil DNA samples, from forest edge and interior sites

Family	Conventional, Edge, specimens	Conventional, Edge, species	Conventional, Interior, specimens	Conventional, Interior, species	Bulk invert DNA, Edge, sequences	Bulk invert DNA, Edge, OTUs	Bulk invert DNA, Interior, sequences	Bulk invert DNA, Interior, OTUs	Soil DNA, Edge, sequences	Soil DNA, Edge, OTUs	Soil DNA, Interior, sequences	Soil DNA, Interior, OTUs
(Coleoptera)					49	1	1	1	16	2	4	2
Anobiidae									6	2	6	2
Anthicidae	1	1			22	1			3	1	2	1
Anthribidae	10	6	1	1								
Cantharidae	1	1	2	1								
Carabidae	130	12	215	11	21125	23	39115	21	215	12	215	20
Cerambycidae	9	6	12	7	35	2	1	1	3	2	7	2
Chrysomelidae	65	4	14	3	3988	12	10683	10	103	8	65	7
Clambidae			1	1								
Cleridae	6	2							2	1		
Coccinellidae	7	2	3	3								
Corylophidae	2	1	1	1					387	12	171	10
Cryptophagidae	1	1	2	1								
Curculionidae	71	20	51	11	6855	14	3441	12	111	3	75	3
Dytiscidae									27	2	5	1
Elateridae	26	3	18	2	7277	10	11003	15	230	6	709	5
Endomychidae									21	1	70	1
Erotylidae	1	1										
Eucinetidae	1	1			6	1						
Eucnemidae	1	1										
Euxestidae	4	1										
Histeridae	1	1										
Hydraenidae					15	1						
Hydrophilidae			16	2			1853	3				
Latridiidae	12	2	18	3	9	1	2	1				
Leiodidae	10	2	1	1	260	1	51	1				
Lucanidae	2	2	2	1								
Melandryidae	13	2	1	1								
Melyridae					916	3	12	2				
Mycetophagidae			1	1								
Nitidulidae	1	1	1	1					698	3	269	3
Oedemeridae	2	1	5	1	3	1	4149	1				

Family	Conventional, Edge, specimens	Conventional, Edge, species	Conventional, Interior, specimens	Conventional, Interior, species	Bulk invert DNA, Edge, sequences	Bulk invert DNA, Edge, OTUs	Bulk invert DNA, Interior, sequences	Bulk invert DNA, Interior, OTUs	Soil DNA, Edge, sequences	Soil DNA, Edge, OTUs	Soil DNA, Interior, sequences	Soil DNA, Interior, OTUs
Phalacridae					4	1						
Ptiliidae									24	2	10	2
Ptinidae									47	1	39	1
Pyrochroidae	4	1							10	1		
Scarabaeidae	17	1	4	2	1789	4	4436	7	43	3	33	2
Scirtidae	1	1	9	1								
Scraptiidae	2	1	1	1								
Silvanidae									1	1	2	1
Staphylinidae	48	12	52	8	21372	6	7334	6	101	6	147	6
Tenebrionidae	4	2			10	2						
Trogidae					10	1						
Ulodidae	2	1	1	1								
Zopheridae	20	3	15	3								

Table S4. OTUs with top BLAST matches of ≥ 97 % identity over ≥ 250 bp to terrestrial invertebrate sequences

OTU	Genbank match ID	Identity (%)	Length (bp)	Phylum	Class	Order	Family	Genus and Species	Substrate
Otu5341	JQ909120	98.4	313	Annelida	-	Haplotaxida	Lumbricidae	*Lumbricus rubellus*	Invert DNA
Otu1458	KT706008	98.72	313	Annelida	-	Haplotaxida	Lumbricidae	*Lumbricus* sp.	Soil DNA
Otu108	JQ909008	100	313	Annelida	-	Haplotaxida	Lumbricidae	*Aporrectodea rosea*	Soil DNA
Otu1725	KT073961	100	313	Annelida	-	Haplotaxida	Lumbricidae	*Octolasion cyaneum*	Soil DNA
Otu668	JQ347512	99.04	313	Arthropoda	Arachnida	Araneae	Clubionidae	*Clubiona peculiaris*	Both
Otu4511	JN377963	99.68	313	Arthropoda	Arachnida	Araneae	Clubionidae	*Clubiona huttoni*	Invert DNA
Otu5307	JN377965	99.68	313	Arthropoda	Arachnida	Araneae	Clubionidae	*Clubiona consensa*	Invert DNA
Otu1202	AY059977	98.4	313	Arthropoda	Arachnida	Araneae	Lycosidae	*Anoteropsis hilaris*	Invert DNA
Otu17594	KF669362	100	313	Arthropoda	Arachnida	Araneae	Thomisidae	*Sidymella angularis*	Both
Otu3247	KR070469	97.75	311	Arthropoda	Arachnida	-	Eupodidae	*Eupodidae* sp.	Both
Otu549	KJ871363	100	313	Arthropoda	Arachnida	Opiliones	Phalangiidae	*Phalangium opilio*	Invert DNA
Otu4298	KT808353	100	313	Arthropoda	Collembola	-	Hypogastruridae	*Hypogastrura purpurescens*	Both
Otu2334	KM617402	99.36	313	Arthropoda	Collembola	-	Isotomidae	*Desoria* sp.	Soil DNA
Otu63	KJ716825	100	286	Arthropoda	Collembola	-	Isotomidae	*Hemisotoma* sp.	Soil DNA
Otu90	JN970927	100	313	Arthropoda	Collembola	-	Neelidae	*Megalothorax* sp.	Soil DNA
Otu370	KM611675	100	313	Arthropoda	Diplopoda	Julida	Julidae	*Ophyiulus pilosus*	Both
Otu173	KF551681	99.26	272	Arthropoda	Insecta	Coleoptera	Carabidae	*Aulacopodus calathoides*	Invert DNA
Otu11768	KJ965508	97.76	313	Arthropoda	Insecta	Coleoptera	Carabidae	*Harpalus affinis*	Invert DNA
Otu2364	JN171146	97.75	311	Arthropoda	Insecta	Coleoptera	Carabidae	*Oopterus helmsi*	Invert DNA
Otu1795	KT957124	100	311	Arthropoda	Insecta	Coleoptera	Chrysomelidae	*Eucolaspis* sp.	Both
Otu1213	KJ418102	100	313	Arthropoda	Insecta	Coleoptera	Curculionidae	*Listronotus bonariensis*	Invert DNA
Otu1657	KM447349	99.36	313	Arthropoda	Insecta	Coleoptera	Curculionidae	*Sitona lepidus*	Invert DNA
Otu9431	HQ453126	99.68	313	Arthropoda	Insecta	Coleoptera	Curculionidae	*Naupactus leucoloma*	Soil DNA
Otu147	JN793484	98.72	313	Arthropoda	Insecta	Coleoptera	Scarabaeidae	*Costelytra zealandica*	Both
Otu526	EU418572	100	313	Arthropoda	Insecta	Diptera	Calliphoridae	*Calliphora vicina*	Invert DNA
Otu16995	EU418563	97.44	313	Arthropoda	Insecta	Diptera	Calliphoridae	*Calliphora stygia*	Invert DNA
Otu860	EU493564	100	313	Arthropoda	Insecta	Diptera	Ephydridae	*Hydrellia tritici*	Both
Otu11815	KR988375	98.71	311	Arthropoda	Insecta	Diptera	Psychodidae	*Psychoda* sp.	Invert DNA
Otu7	GQ254423	100	313	Arthropoda	Insecta	Diptera	Sarcophagidae	*Oxysarcodexia varia*	Both
Otu15194	GQ254423	98.67	300	Arthropoda	Insecta	Diptera	Sarcophagidae	*Oxysarcodexia varia*	Invert DNA
Otu10143	JX418167	99.04	313	Arthropoda	Insecta	Diptera	Sciaridae	*Bradysia pallipes*	Invert DNA
Otu8068	JQ613826	100	313	Arthropoda	Insecta	Diptera	Sciaridae	*Lycoriella castanescens*	Invert DNA
Otu4596	KR039842	99.68	313	Arthropoda	Insecta	Hemiptera	Aphrophoridae	*Philaenus spumarius*	Invert DNA
Otu8717	JQ240196	99.67	303	Arthropoda	Insecta	Hemiptera	Nabidae	*Nabis kinbergii*	Invert DNA
Otu2720	JX675425	100	310	Arthropoda	Insecta	Hemiptera	Cicadidae	*Amphipsalta zelandica*	Soil DNA
Otu966	AF493568	99.04	313	Arthropoda	Insecta	Hemiptera	Psyllidae	*Acizzia uncatoides*	Soil DNA

OTU	Genbank match ID	Identity (%)	Length (bp)	Phylum	Class	Order	Family	Genus and Species	Substrate
Otu1623	JQ843647	100	313	Arthropoda	Insecta	Hymenoptera	Apidae	*Bombus terrestris*	Invert DNA
Otu2898	KJ734255	100	313	Arthropoda	Insecta	Hymenoptera	Apidae	*Bombus ruderatus*	Invert DNA
Otu8624	AY181102	100	313	Arthropoda	Insecta	Hymenoptera	Apidae	*Bombus hortorum*	Invert DNA
Otu6367	AY427879	100	313	Arthropoda	Insecta	Hymenoptera	Braconidae	*Aphidius ervi*	Invert DNA
Otu502	FJ824424	98.4	313	Arthropoda	Insecta	Hymenoptera	Formicidae	*Huberia striata*	Both
Otu3619	KF391932	99.66	293	Arthropoda	Insecta	Lepidoptera	Geometridae	*Chloroclystis filata*	Invert DNA
Otu5750	KX047005	100	313	Arthropoda	Insecta	Lepidoptera	Noctuidae	*Agrotis ipsilon*	Invert DNA
Otu525	JF818797	99.04	313	Arthropoda	Insecta	Lepidoptera	Oecophoridae	*Tingena armigerella*	Invert DNA
Otu1	KF491658	100	313	Arthropoda	Insecta	Lepidoptera	Tortricidae	*Cryptaspasma querula*	Both
Otu11828	KF491658	97.12	312	Arthropoda	Insecta	Lepidoptera	Tortricidae	*Cryptaspasma querula*	Invert DNA
Otu32	FJ225492	100	313	Arthropoda	Insecta	Lepidoptera	Tortricidae	*Ctenopseustis obliquana*	Both
Otu3035	FJ225540	99.04	312	Arthropoda	Insecta	Lepidoptera	Tortricidae	*Ctenopseustis fraterna*	Invert DNA
Otu170	FJ225650	99.36	313	Arthropoda	Insecta	Lepidoptera	Tortricidae	*Planotortrix octo*	Invert DNA
Otu2421	JQ240180	97.76	313	Arthropoda	Insecta	Neuroptera	Hemerobiidae	*Micromus tasmaniae*	Invert DNA
Otu2163	JF895573	97.12	313	Arthropoda	Insecta	Orthoptera	Anostostomatidae	*Hemiandrus pallitarsis*	Invert DNA
Otu4351	JF895571	97.76	313	Arthropoda	Insecta	Orthoptera	Anostostomatidae	*Hemiandrus pallitarsis*	Invert DNA
Otu7945	JF895557	98.4	313	Arthropoda	Insecta	Orthoptera	Anostostomatidae	*Hemiandrus pallitarsis*	Invert DNA
Otu6838	JN409958	99.68	313	Arthropoda	Insecta	Orthoptera	Rhaphidophoridae	*Talitropsis sedilloti*	Invert DNA
Otu118	KM531703	100	291	Arthropoda	Insecta	Psocoptera	-	*Psocoptera* sp.	Both
Otu2874	KX291952	99.36	313	Arthropoda	Insecta	Trichoptera	Hydroptilidae	*Hydroptilidae* sp.	Invert DNA
Otu12380	KF894378	100	306	Mollusca	Gastropoda	-	Agriolimacidae	*Deroceras reticulatum*	Invert DNA
Otu528	KM491205	98.71	310	Nematoda	Chromadorea	Tylenchida	Meloidogynidae	*Meloidogyne hapla*	Soil DNA

Table S5. Results of Procrustes and Mantel test comparisons of overall community structure based on invertebrate specimens, bulk invertebrate DNA, and soil DNA from forest edge and interior samples, and the beetle components of these datasets

Test	Datasets compared	Statistic	Significance
Procustes test	specimens vs. bulk invertebrate DNA	0.552	0.054
	specimens vs. soil DNA	0.327	0.569
	bulk invertebrate DNA vs. soil DNA	0.523	0.105
	beetle specimens vs. bulk DNA beetles	0.772	< 0.001
	beetle specimens vs. soil DNA beetles	0.469	0.213
	bulk DNA beetles vs. soil DNA beetles	0.407	0.378
Mantel test	specimens vs. bulk invertebrate DNA	0.226	0.082
	specimens vs. soil DNA	-0.003	0.455
	bulk invertebrate DNA vs. soil DNA	0.171	0.155
	beetle specimens vs. bulk DNA beetles	0.353	0.004
	beetle specimens vs. soil DNA beetles	0.05	0.37
	bulk DNA beetles vs. soil DNA beetles	-0.012	0.513

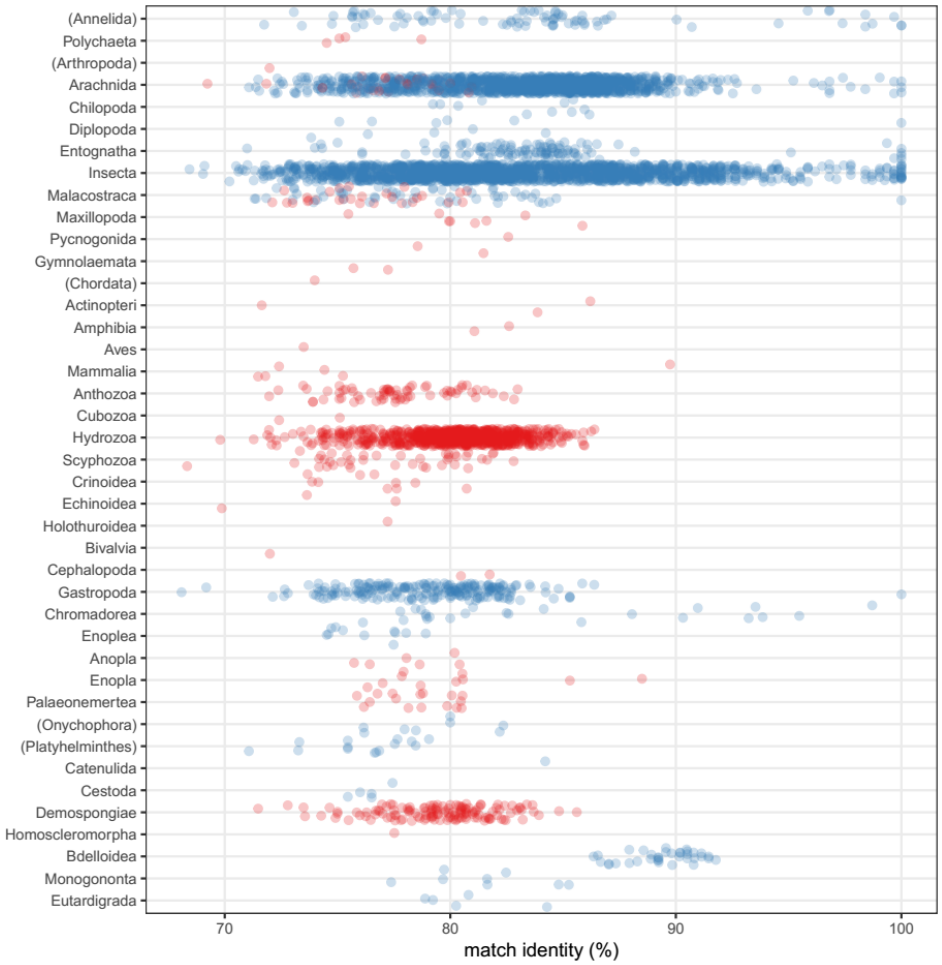


Figure S3. Taxonomic Class and pairwise sequence identity (%) of the top BLAST matches to bulk invertebrate DNA and soil DNA OTUs. Blue and red data points indicate matches to sequences from expected terrestrial invertebrates and matches to sequences from non-terrestrial or non-invertebrate organisms, respectively.

Table S6. Comparison of conventional sampling versus DNA metabarcoding for monitoring invertebrate communities. Costs are shown as a percentage of total costs. Data quality and application assesses the advantages (+) and disadvantages (-) of conventional sampling and DNA metabarcoding methods

			DNA metabarcoding	
		Conventional sampling (malaise & pitfall traps)	**Bulk invertebrate DNA** (malaise & pitfall traps)	**Soil DNA**
COSTS	Sampling in the field	15	15	5
	Laboratory Analysis: Counting and morphological identifications	70	NA	NA
	Laboratory Analysis: DNA extraction and sequencing	NA	30	35
	Bioinformatics	NA	40	45
	Data analyses	15	15	15
DATA QUALITY AND APPLICATION	Quantitative	Yes; counting of specimens (+)	Partially; restricted by PCR complications (-)	
	Taxonomic coverage	Limited; large volume of specimens in samples (-)	Very wide (+)	
	Taxonomic depth	Restricted by scarcity of taxonomic expertise in some groups (-)	Restricted by database coverage (-)	
	Interior/edge ecological trends detected	Yes (+)	Yes (+)	Less confident (-)
	Expertise required	Taxonomic knowledge (-)	Bioinformatics skills(-)	

	Conventional sampling (malaise & pitfall traps)	DNA metabarcoding	
		Bulk invertebrate DNA (malaise & pitfall traps)	Soil DNA
Interpretation	Easy (+)	More challenging, due to methodological uncertainties and massive data (-)	
Specimens retained	Yes (+)	Depending on approach (e.g. take a leg for DNA metabarcoding and leave rest for specimen identification)	NA
Application for end-users	Easy (+)	Promising but requires improvement (-)	

References

Edgar R (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature Methods 10: 996-998

Martin M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17: 10-12 DOI: http://dx.doi.org/10.14806/ej.17.1.200

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ. 4: e2584