Austral Entomology

Austral Entomology (2019) **58**, 675–686

DNA metabarcoding as a 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.*

Abstract When conserving 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 (pitfall and malaise traps) and a DNA metabarcoding approach applied to two alternative sample types (conventional invertebrate samples and soil samples). The bulk invertebrate and soil DNA metabarcoding methods were 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. DNA metabarcoding offers conservation managers a practical, cost-effective technique for characterising whole invertebrate communities. However, increasing the taxonomic coverage of reference sequence databases (particularly for New Zealand invertebrates) through DNA barcoding efforts should be the focus of future research as it would improve the utility of metabarcoding methods for invertebrate monitoring, which would complement conventional techniques.

Key words COI, edge effect, environmental DNA, forest.

INTRODUCTION

Invertebrates constitute a substantial proportion of biodiversity and are critical to ecosystem functions (Lavelle *et al.* 2006; Watts *et al.* 2008). However, their inclusion in biodiversity monitoring and conservation planning lags behind more widely appreciated taxa (Gerlach *et al.* 2013) and is usually limited to large iconic invertebrates such as carabid beetles in Europe (Rainio & Niemela 2003) or wētā in New Zealand (Watts *et al.* 2017). A reason for this is because collecting and sorting invertebrates using conventional monitoring techniques is often expensive, time consuming and restricted by expertise in diagnostics.

Environmental DNA metabarcoding techniques can characterise entire invertebrate communities from a single environmental sample (Bohmann *et al.* 2014; Holdaway *et al.* 2017). DNA metabarcoding analyses involve extraction of DNA from environmental samples such as bulk invertebrates, soil, leaf litter or water and then amplification and sequencing of specific gene regions, grouping of sequences into OTUs based on similarity and identification of OTUs by comparing their sequences to available reference databases (e.g. Yu *et al.* 2012; Bohmann *et al.* 2014; Drummond *et al.* 2015; Holdaway *et al.* 2017). DNA metabarcoding analysis of bulk invertebrate samples (such as malaise trap collections) circumvents time-consuming, costly and demanding microscope-based identifications (Yu *et al.* 2012; Morinière *et al.* 2016). Soil is a promising substrate for invertebrate DNA metabarcoding analyses, as it is easily sampled and contains DNA from both soil-dwelling invertebrates (e.g. mites, nematodes, earthworms and insect larvae) and above-ground organisms (e.g. as carcases, exoskeletons and frass; Drummond *et al.* 2015). However, the presence of relic DNA (Carini *et al.* 2016) along with a myriad of microorganisms makes the interpretation of soil DNA metabarcoding results more challenging than bulk invertebrate material.

DNA metabarcoding could be an important monitoring tool for invertebrates within many conservation projects. However, we need to apply both DNA metabarcoding and conventional methods to the same taxonomic groups to demonstrate and improve the accuracy and reliability of DNA-based methods and to assess the additional value (Ji *et al.* 2013; Holdaway *et al.* 2017). The *Cape-to-City* project (Hawke's Bay, New Zealand) aims to restore indigenous biodiversity across 26 000 ha of farmed landscape, through mammal predator

*wattsc@landcareresearch.co.nz

© 2019 Australian Entomological Society

doi: 10.1111/aen.12384

control and habitat restoration (Norbury & McLennan 2015). Using this 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. An identical budget (NZ$20 000 in 2015) for each approach enabled direct cost–benefit comparison. We collected and identified invertebrate specimens using conventional methods (malaise and pitfall traps) and then used DNA metabarcoding methods to analyse two different sample types: (1) DNA extracted from the bulk conventional invertebrate samples and (2) DNA extracted from soil collected at the same locations. This design enabled comparison of soil and above-ground (pitfall and malaise trap) communities, as well as between conventional monitoring and DNA metabarcoding.

We focused on differences in invertebrate communities on the forest edge vs. the interior to detect fragmentation effects, in particular on the sensitive and well-described beetle fauna (Insecta: Coleoptera) (Didham *et al*. 1998; Ewers & Didham 2008). We hypothesised that metabarcoding analysis of DNA would, for a similar cost, (1) result in the detection of a broader range of taxa and increased taxonomic resolution from bulk invertebrate samples compared with conventional identification methods; (2) reproduce patterns of diversity and composition detected by conventional methods, and variation in these patterns between forest interior and edge; and (3) demonstrate that soil DNA extracts would show similar community patterns to the conventional trap samples.

MATERIALS AND METHODS

Study area and sampling design

The study site (Mohi Bush Scenic Reserve: 61 ha; 39.85643S, 1769011E) is 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 a canopy dominated by tawa (*Beilschmiedia tawa*), with scattered emergent podocarps (miro (*Prumnopitys ferruginea*) and matai (*Prumnopitys taxifolia*)) overtopping a diverse subcanopy layer including the trees *Hedycarya arborea* and *Melicytus ramiflorus*.

Twelve 20 × 20 m plots were located in the 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). Half of each edge plot was in pasture and half in forest. For supplementary information on the study area and sampling design, refer to Appendix S1.

Conventional invertebrate collection and identification

Invertebrates were sampled using a modified malaise trap (Watts *et al*. 2012) and pitfall traps placed in the centre of each 20 × 20 m plot (Fig. S2). At the edge plots, the malaise trap was located in the forest (<3 m from the pasture-forest edge). The malaise trap collecting jar 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 5 m 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 100% ethanol.

Invertebrates were sorted and counted to order level using a binocular microscope. Beetles were sorted to recognised taxonomic units (hereafter referred to as species) and, where possible, given generic and species-level identifications by an expert diagnostician (Stephen Thorpe, Independent diagnostic researcher). After sorting and counting, the specimens from the malaise trap and four pitfall traps from each plot were pooled into a single bulk invertebrate sample for DNA analyses.

Soil sample collection

Soil samples were collected from the same twelve 20 × 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 per plot with samples randomly located following standard sampling protocols (Fig. S2) using sterilised trowels. The soil samples were stored at 4° 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 300 mg of the material using the Machery-Nagel NucleoSpin 96 Tissue extraction kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany). Further details of the invertebrate homogenisation and DNA extraction protocols are described in Appendix S2.

The soil samples from each plot were homogenised, and a 10 g subsample was taken from which DNA was extracted using the PowerMax Soil DNA Isolation Kit (Qiagen Sciences, Germantown, MD, USA), following the manufacturer's 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 linker sequences attached to their 5′ ends to enable the subsequent addition of sequencing adaptors (Clark *et al*. 2014). These primers result in a 313 bp amplicon, corresponding to the 5′ end of the widely used COI DNA barcode region (Folmer *et al*. 1994). The initial template amplification stage was carried out using a Touchdown protocol (Leray *et al*. 2013) as follows: initial denaturation at 95°C for 10 m; 16 cycles of 95°C for 10 s, 62°C to 47°C (reducing by 1 each cycle) for 30 s, 72°C for 60 s; then 25 cycles of 95°C for 10 s, 46°C for 30 s, 72°C for 60 s; 72°C for 10 min. This was followed by the second stage of the MoTasp protocol (Clark *et al*. 2014) to add the sample-specific barcodes and sequencing adaptors, as follows: 95°C for 10 m; 40 cycles of 92°C for 15 s, 54°C for 30 s, 72°C for 30 s; 72°C for 10 m. Further details of the PCR protocols are described in Appendix S2. The amplicons were sequenced by New Zealand Genomics Limited in a 2 × 250 bp Illumina MiSeq run.

© 2019 Australian Entomological Society

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 using fastq_mergepairs in USEARCH v9 (Edgar 2013). Amplicon primers were trimmed from the merged sequences using cutadapt v1.11 (Martin 2011). The merged sequences were quality filtered (fastq_filter) with maximum expected errors (fastq_maxee) of 1.0 and dereplicated (derep_fulllength) with a minimum sequence length threshold (minseqlength) of 200 bp, in VSEARCH (Rognes *et al*. 2016). The dereplicated sequences were filtered for chimeras and clustered into OTUs at a 97% identity threshold with a minimum abundance threshold of 2 sequences (minsize 2) in a single step using the UPARSE algorithm (cluster_otus) within USEARCH v9 (Edgar 2013). Finally, the 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 using the RDP Naïve Bayesian classifier (Wang *et al*. 2007) in combination with the RDP-formatted MIDORI UNIQUE 1.1 mitochondrial COI database (Machida *et al*. 2017), which contains metazoan COI sequences with species-level identifications derived from GenBank. We modified this database by including bacterial, fungal and protist COI sequences with species-level identifications extracted from GenBank, to enable the detection of non-metazoan OTUs in our data.

Only OTUs assigned to an expected terrestrial invertebrate phylum were included in biodiversity analyses. OTUs were retained if the taxonomic identity 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). The Annelida, Malacostraca and Gastropoda OTUs were further filtered and retained only if the identity was not one of the following non-terrestrial groups: Bivalvia, Cephalopoda, Decapoda, Mysida, Nudibranchia or Polychaeta. In addition, for analyses of OTUs at taxonomic levels of order and below, we used minimum bootstrap support cut-offs to limit the data to OTUs with likely correct assignments at each rank (5% for order, 25% for family, 60% for genus and 90% for species) based on Porter and Hajibabaei (2018). For the family-level analysis of beetles, however, we applied a family-level cut-off of 5%, as a 25% cut-off resulted in a failure to detect most beetle families known to be present in the bulk invertebrate samples. Any sequence reads occurring in the negative control samples were subtracted from their corresponding OTUs in all other samples.

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 identifications (i.e. how many OTUs could be identified to species, genus, family and order). We compared this with the taxonomic composition, resolution and abundance of the same invertebrate samples determined using conventional methods, to assess whether the same taxa were detected by each method. Additionally, we investigated whether the specimen and sequence abundances detected by each method were correlated. We also compared the taxonomic composition of the soil DNA dataset to that of the conventional and bulk invertebrate DNA results, to determine whether different sampling approaches resulted in the detection of different ranges of taxa.

The mean numbers of specimens, sequences or OTUs detected in different taxonomic groups 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). We used non-metric multidimensional scaling ordinations based on presence/absence Jaccard distance for DNA metabarcoding data and quantitative Bray–Curtis dissimilarity 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 specimens, bulk invertebrate DNA OTUs and soil DNA invertebrate OTUs) 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 7503 (4069 from the forest edge and 3434 from the forest interior) invertebrate specimens were collected using conventional sampling methods (Table S1). Arthropoda was the most common phylum (>99% of specimens), followed by Platyhelminthes and Mollusca. The Arthropoda specimens included five classes and ≥20 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). Opiliones, Diptera and Orthoptera specimen abundances were 5-fold, 1.6-fold and 2.8-fold higher in the edge samples than in the interior samples, respectively, according to two-sided Student's *t*-tests (t = 4.41, 5.56 and 4.11; P_{adj} = 0.036, 0.007 and 0.036; df = 10). Hymenoptera specimens were fourfold more abundant in edge samples than interior samples with marginal significance (t = 3.65; P_{adj} = 0.063; df = 10). Conversely, Collembola specimen abundance was 2.5-fold higher in the interior samples than in the edge samples (t = −10.31; $P_{adj} < 0.001$; df = 10).

A total of 34 beetle families, including 108 genera and 121 species, were detected, Carabidae being most abundant (345 specimens) followed by Curculionidae (122 specimens), Staphylinidae (100 specimens), Chrysomelidae (79 specimens)

© 2019 Australian Entomological Society

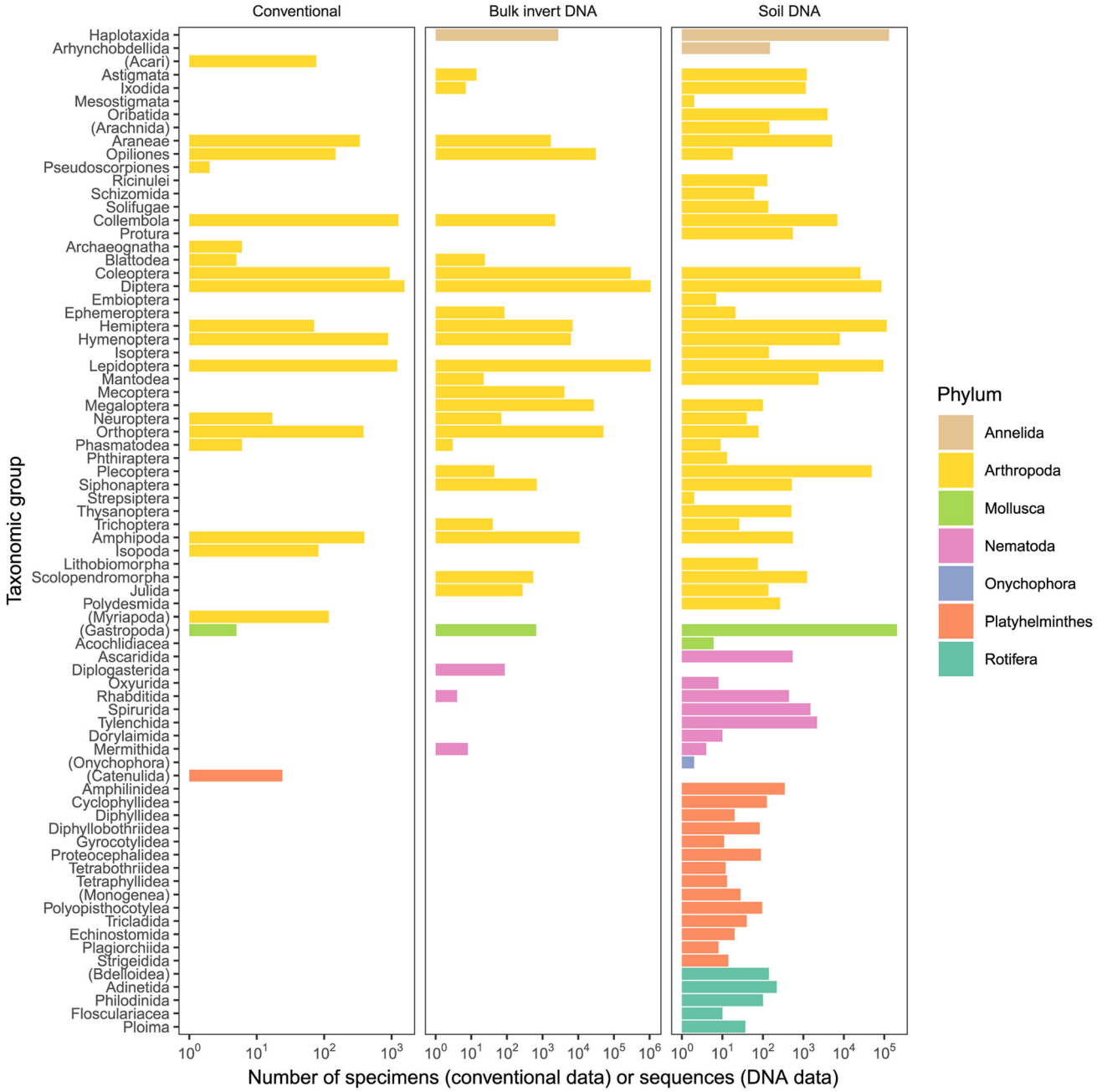

Fig. 1. Overall taxonomic composition of invertebrate specimens and DNA sequence reads derived from DNA metabarcoding analysis of bulk invertebrate DNA and soil DNA extracts. Taxonomic groups are orders, except where indicated by parentheses.

and Elateridae (44 specimens; 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 vs. interior samples.

Overall DNA sequencing and OTU identification results

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, resulting in a total of 5061 OTUs. Of these, 2154 OTUs (3.3 million sequences) were assigned to an expected terrestrial invertebrate phylum (Fig. S3). The other 2907 OTUs were assigned to non-terrestrial or non-invertebrate groups, including Ascomycota, Stramenopiles, Cnidaria and Porifera, and were excluded from further analysis. Almost all of these OTUs (2796) occurred only in the soil DNA results. Of the OTUs matching an expected terrestrial invertebrate phylum, 650 occurred only in bulk invertebrate samples, 1342 occurred only in soil samples and 162 occurred in both sample types. Of these, 2016 OTUs were successfully assigned to a taxonomic order, 415 to a family and 110 to a genus. Only 72 OTUs

© 2019 Australian Entomological Society

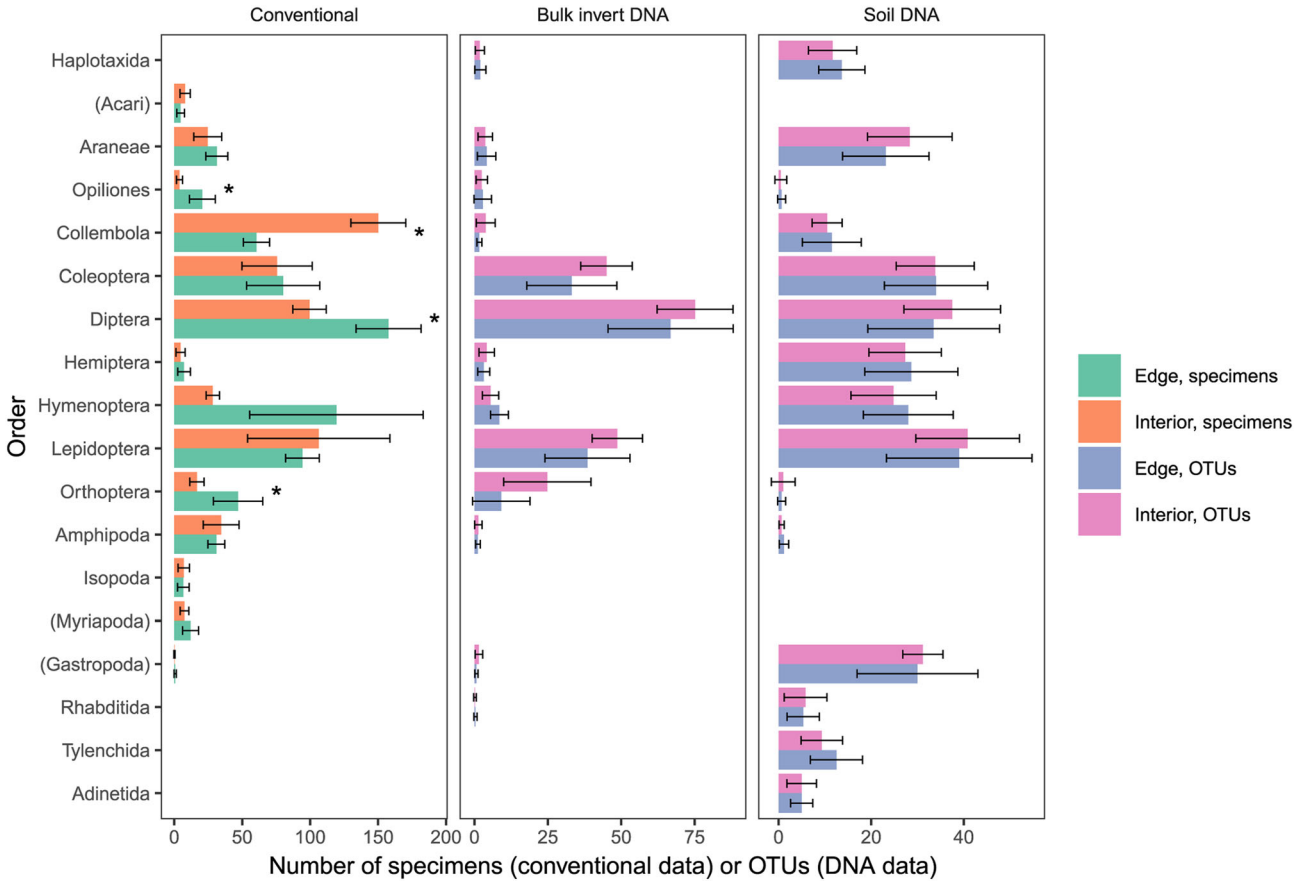

Fig. 2. Mean numbers of invertebrate specimens detected per order by conventional methods and invertebrate OTUs detected per order by DNA metabarcoding methods, in forest edge and interior samples. Only orders represented by at least five mean specimens or OTUs per sample are included. Error bars represent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

were identified to species level, based on a species-level identification bootstrap confidence score >90% (Table S4). These included 44 Insecta OTUs, 10 Annelida OTUs, 2 Collembola OTUs, 12 Arachnida OTUs, a single OTU from each of Diplopoda, Gastropoda, Malacostraca and Chromadorea and a total of 47 genera. The most abundant of these OTUs (>600 000 sequences) was identified as *Cryptaspasma querula* (Tortricidae). There were eight species to which more than one OTU was assigned with confidence >90%, including nine OTUs assigned to *Hemiandrus pallitarsus*.

Composition of bulk invertebrate DNA results compared to conventional results

The metabarcoding analysis of DNA from the bulk invertebrate specimens resulted in the detection of sequences and OTUs from four phyla, among which Arthropoda was dominant (99.9% of sequences, 98% of OTUs), with small numbers of Annelida, Mollusca and Nematoda (Fig. 1; Table S2). Unlike the conventional specimen results, Platyhelminthes were not detected. Five classes and 25 orders of Arthropoda were detected among the bulk invertebrate DNA results. Fourteen of these orders matched those detected by conventional specimen identifications, and another four corresponded to specimens identified by conventional methods only to class (Acari) or subphylum (Myriapoda) levels. A further seven orders of Arthropoda detected among the bulk invertebrate DNA results were not detected by the conventional methods (Archaeognatha, Ephemeroptera, Mecoptera, Megaloptera, Siphonaptera, Thysanoptera and Trichoptera). Conversely, three Arthropoda orders detected by the conventionally methods were absent from the bulk invertebrate DNA results (Pseudoscorpiones, Archaeognatha and Isopoda).

Three of the four most abundant orders based on conventional identifications were also the most abundant based on sequence counts (Diptera, Lepidoptera and Coleoptera, but not Collembola), and these were also the most species-rich (OTUs) groups. There was a moderate correlation between raw sequence abundances and specimen abundances for orders that were detected by both the conventional method and the bulk invertebrate DNA sequencing method in each sample (Pearson correlation = 0.43, $P < 0.0001$, $df = 143$). Stronger correlations were observed when sequence counts were square root transformed (Pearson correlation = 0.50, $P < 0.0001$, $df = 143$) and when specimen and sequence counts were summed by order across all samples (Pearson correlation = 0.70, $P = 0.001$, $df = 15$) or were aggregated to class level (Pearson correlation = 0.85, $P < 0.0001$, $df = 50$). The observed differences in specimen abundances between edge and interior samples based on

© 2019 Australian Entomological Society

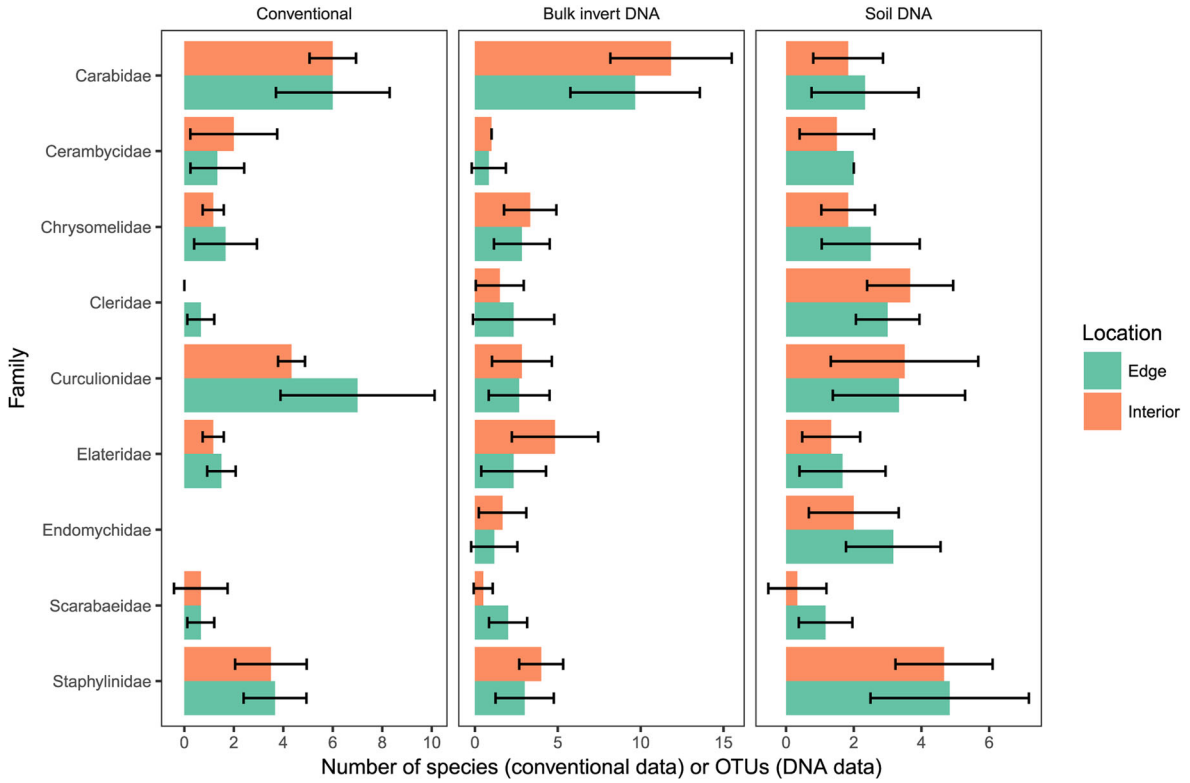

Fig. 3. Mean numbers of beetle species and OTUs per family detected in forest edge and interior samples using conventional invertebrate sampling and DNA metabarcoding methods. Only families represented by at least two mean species or OTUs per sample are included. Error bars represent ±95% CI.

conventional identifications were not clearly reflected in the bulk invertebrate DNA sequence abundances, with no orders having significantly differing sequence or OTU counts between edge and interior samples (Fig. 2).

Eighteen beetle families were detected among the bulk invertebrate DNA results, compared to 34 detected among the conventional beetle specimen identifications, and only 13 beetle families matched between both datasets. Five beetle families detected in bulk invertebrate DNA results were not detected by conventional specimen identifications, whereas the opposite was true of 21 families. Two of the conventionally identified beetle families were unrepresented in the sequence identification database used, precluding their identification among the DNA results. Only nine of 108 beetle genera identified in the conventional results were also successfully detected in the bulk invertebrate DNA results. However, most of the conventionally identified beetle genera (67%) were unrepresented in the sequence identification database.

The total number of beetle OTUs (152) exceeded the number of conventionally identified beetle species (121), although the number of beetle OTUs assigned to a family (119, with a total of 93 646 sequences) was very similar to the number of species (Table S3). The most abundant beetle family in the bulk invertebrate DNA results was Carabidae, as for the conventional beetle results. The next four most abundant beetle families (Elateridae, Chrysomelidae, Staphylinidae and Curculionidae) also matched those in the conventional data but in a different order. The number of beetle OTUs in each of these families (except for Curculionidae) exceeded the numbers of species identified by conventional means (Fig. 3; Table S3). No beetle families had significantly differing numbers of sequences or OTUs between edge and interior samples (Fig. 3).

We observed a modest correlation between beetle specimen abundances multiplied by length (as a proxy for biomass) and square root-transformed sequence abundances in different samples for the nine detected genera (Pearson correlation = 0.504, $P = 0.01009$; Fig. S4).

Composition of soil DNA results

The metabarcoding analysis of soil DNA detected a broader range of taxa (seven phyla and >65 orders) than the conventional specimens or bulk invertebrate DNA (Figs 1,2; Table S2). Arthropoda were again most abundant (55% of sequences, 72% of OTUs), followed by Mollusca (27% of sequences, 9% of OTUs), Annelida (17% of sequences, 4% of OTUs), Nematoda (0.7% of sequences, 7% of OTUs) and Platyhelminthes, Rotifera and Onychophora (all <0.2% of sequences, ≤4% of OTUs). Compared with the bulk invertebrate DNA results, five additional hexapod groups were detected in the soil DNA results (Embioptera, Isoptera, Phthiraptera, Strepsiptera and Protura), as were five additional orders of Arachnida (Mesostigmata, Oribatida, Ricinulei, Schizomida and Solifugae), two additional orders of Myriapoda (Lithobiomorpha and Polydesmida), five additional orders of Nematoda and one additional Annelida order (Arhynchobdellida), along with 14 orders of Platyhelminthes

© 2019 Australian Entomological Society

and five orders of Rotifera. Of the orders detected by the conventional methods, Pseudoscorpiones, Archaeognatha and Isopoda were absent from the soil DNA results (as they were from the bulk invertebrate DNA results). Just two orders detected in the bulk invertebrate DNA results were absent from the soil DNA results (Blattodea and Diplogasterida). The most abundant orders of Arthropoda in the soil DNA results were Hemiptera, Diptera, Lepidoptera, Coleoptera and Plecoptera. As for the bulk invertebrate DNA results, there were no significant differences in the mean abundances of soil DNA sequences or OTUs per order in the edge compared to interior samples.

A total of 129 beetle OTUs were identified from the soil DNA, of which 65 OTUs (with a total of 7126 sequences) were assigned to 13 beetle families. There were 23 and seven beetle families present among the conventional and bulk invertebrate DNA results respectively that were not detected in the soil DNA, whereas there were two beetle families in the soil DNA results that were not detected in the conventional results and another two not detected in the bulk invertebrate DNA results (Table S3). The most abundant beetle families in the soil DNA results were not consistent with those in the bulk invertebrate DNA results or the conventional specimens. Curculionidae were most abundant, followed by Buprestidae, Cerambycidae, Staphylinidae and Cleridae. The beetle families with the most OTUs were Staphylinidae, Curculionidae, Endomychidae, Carabidae and Cleridae, although there were fewer OTUs in each of these families than in the conventional or bulk invertebrate DNA results. As for the bulk invertebrate DNA results, no beetle families had significantly differing numbers of sequences or OTUs between edge and interior samples (Fig. 3).

Community structure similarities between datasets

Multivariate ordinations of sample similarity indicated that the composition of overall invertebrate communities and beetle communities detected in forest edge samples differed from communities detected in forest interior samples, according to both DNA metabarcoding approaches, as did beetle communities according to all sampling methods (Fig. 4). In contrast, the overall invertebrate composition according to conventional methods did not differ between forest edge and interior samples, except for when an abundance-weighted distance metric was used (whereas interior–edge patterns were less clear for the DNA metabarcoding data using an abundance-weighted metric) (Figs S5,S6). Procrustes and Mantel tests indicated that the invertebrate communities detected by metabarcoding analysis of bulk invertebrate samples and soil samples had similar patterns of community structure, as did the beetle communities detected by conventional methods and metabarcoding of bulk invertebrate samples and to a lesser extent, metabarcoding of soil samples. In contrast, there was no evidence of similar community structure between the overall invertebrate composition according to conventional methods and DNA metabarcoding methods.

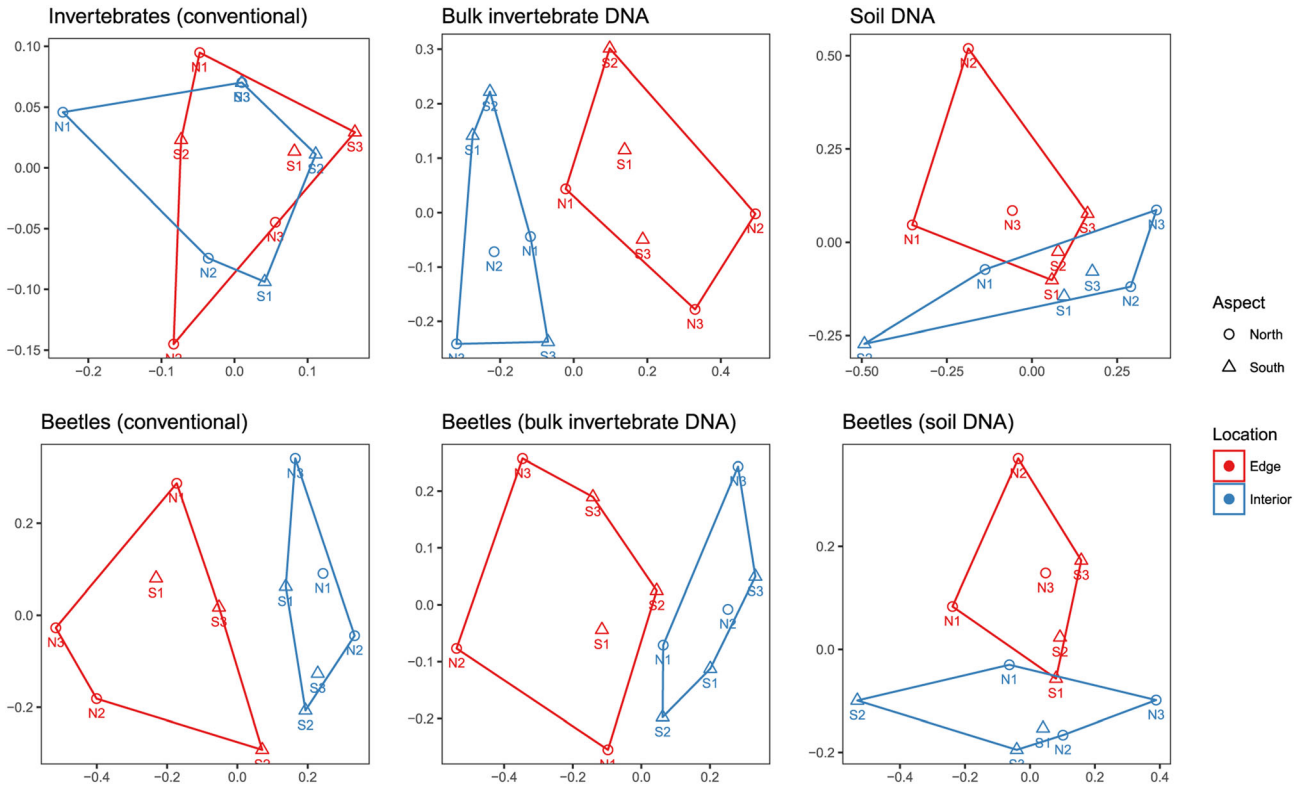

Fig. 4. Non-metric MDS ordinations based on invertebrate specimens, bulk invertebrate DNA and soil DNA metabarcoding results from forest edge and interior samples, and the beetle components of these datasets, using the Jaccard distance metric.

© 2019 Australian Entomological Society

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 (Fig. 5). The clearest discrimination of interior and edge samples was observed for Diptera, Araneae/Opiliones and Lepidotera OTUs detected in the bulk invertebrate DNA dataset. The most overlap between interior and edge samples was observed for Rotifera, Annelida, Hymenoptera and Araneae/Opiliones OTUs detected in the soil DNA dataset (Fig. 5).

DISCUSSION

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

As predicted, DNA metabarcoding methods provided information on a diverse array of invertebrate taxa – more than conventional analysis for the same cost, especially for the soil DNA analysis. However, the sampling method had strong effects on the communities detected, indicating that soil samples tend to recover different components of biodiversity compared to aboveground traps. Our study illustrates the potential for DNA metabarcoding methods to enable the efficient characterisation of whole invertebrate communities, supporting previous metabarcoding studies of bulk invertebrate samples (Yu *et al.* 2012; Morinière *et al.* 2016) or soil samples (Drummond *et al.* 2015), whilst also demonstrating some shortcomings of these techniques.

The bulk invertebrate DNA analysis did not detect all the groups identified by the conventional methods. Three of the four missing groups (Pseudoscorpiones, Archenognatha and Isopoda) were also absent from the soil DNA results, suggesting that this may be due to PCR primer mismatches. Conversely, the bulk invertebrate DNA results included several additional arthropod groups that were not detected by the conventional methods. These likely represent fragments of invertebrate material (or DNA molecules) undetected or unidentifiable by visual means, including parasites (e.g. nematodes) and gut contents, or misidentified specimens. Inconclusive or incorrect OTU identifications due to reference database coverage limitations

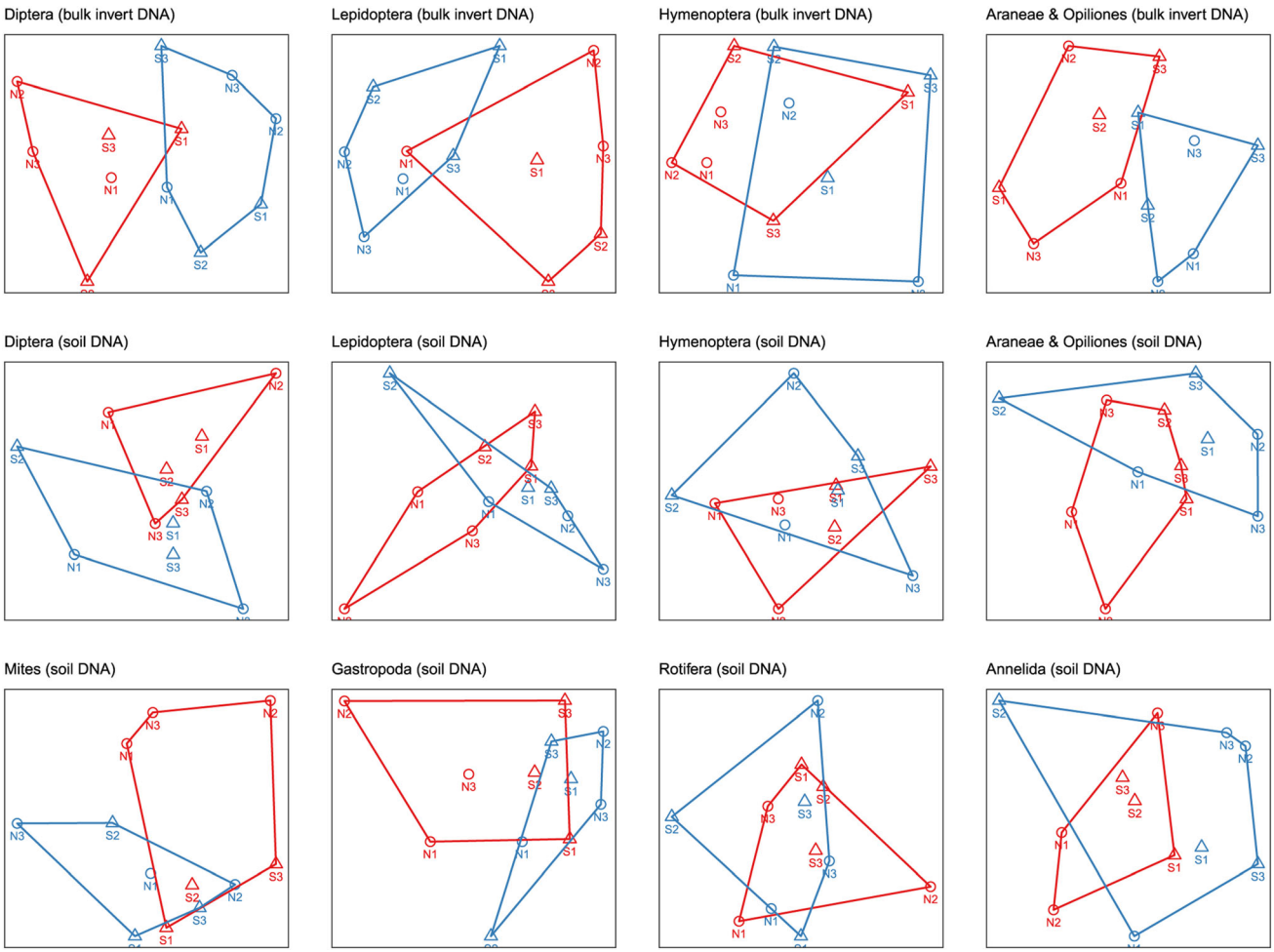

Fig. 5. Non-metric MDS ordinations based on various components of the bulk invertebrate DNA and soil DNA metabarcoding datasets from forest edge and interior samples, using the Jaccard distance metric. Symbols as in Figure 4.

© 2019 Australian Entomological Society

may also have contributed to these taxonomic differences (Kwong *et al.* 2012; Kvist 2013). This factor is highlighted by our beetle DNA metabarcoding results, in which recommended identification thresholds resulted in a failure to detect most beetle families known to be present, likely because most beetle genera and species had no matching reference DNA sequences.

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 soil DNA results included soil-dwelling groups such as mites, nematodes and rotifers, which were largely absent from the conventional and bulk invertebrate DNA results, providing an additional layer of ecological information.

The broader ranges of taxa detected by the DNA metabarcoding analysis methods (especially the soil DNA analysis) are also partly due to the DNA-based identification of certain groups to lower taxonomic ranks than was feasible using conventional methods. For example, mite and myriapod specimens were not identified beyond the ranks of subclass (Acari) or subphylum (Myriapoda) by the conventional analysis, whereas these groups were each represented by up to four different orders in the DNA metabarcoding analysis results. DNA-based methods can thus provide greater taxonomic depth and resolution than can be achieved by conventional identification methods, but this is contingent upon reference DNA sequences matching the groups in question.

The observed ecological patterns (i.e. separation of communities between forest edge and forest interior; Ewers & Didham 2008) in the bulk invertebrate DNA results differed from those obtained through conventional approaches. This may reflect that the DNA metabarcoding approach enabled determination of invertebrate OTUs, approximating species, whereas the conventional invertebrate data (with the exception of beetles) were limited to specimen counts at order level. The bulk invertebrate DNA data also had a broader taxonomic coverage, showing edge effects for a diverse range of invertebrate taxa. Some of the soil invertebrate groups did not show a strong separation of edge vs. interior samples, suggesting that they may be driven by different factors to the above-ground community (Wardle *et al.* 2004 and references therein). In addition, the presence of relic DNA in soils may have prevented the detection of edge effects based on soil DNA (Carini *et al.* 2016).

Our results show that current DNA metabarcoding methods are not a perfect substitute for conventional invertebrate specimen identifications but nevertheless have considerable potential for comprehensive analyses of invertebrate biodiversity based on bulk invertebrate material (Yu *et al.* 2012; Ji *et al.* 2013; Morinière *et al.* 2016) or soil (Drummond *et al.* 2015). Future improvements in DNA sequencing read lengths and reference database coverage should enable more confident sequence identifications, resulting in increased utility of DNA metabarcoding techniques.

Table 1 Comparison of conventional sampling vs. DNA metabarcoding for monitoring invertebrate communities

		Conventional sampling (malaise and pitfall traps)	DNA metabarcoding	
			Bulk invertebrate DNA (malaise and 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 (from abundance data) (+)	Yes (+)	Less confident (−)
	Expertise required	Taxonomic knowledge (−)	Bioinformatics skills (−)	
	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 (−)	

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.

© 2019 Australian Entomological Society

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

Specimen sorting and morphological identifications dominated the costs of conventional invertebrate sampling, whereas the main costs of DNA metabarcoding methods were molecular laboratory procedures and bioinformatic processing (Table 1). Soil is easy to sample in the field (one-off visit), whereas malaise or pitfall traps require specialised sampling and repeat visits.

Conventional methods reliably allow assignment of invertebrates to high-level taxa, but genus and species delineations and identifications are difficult and time consuming. Furthermore, taxonomic identification expertise tends to be scarce and limited to particular taxa. Conventional sampling provides specimens for future research (e.g. taxonomic revision of groups) and readily interpretable abundance data (Table 1). DNA metabarcoding, in contrast, allows for the rapid determination of OTUs (roughly analogous to species or genera) and is effective for efficient detection of a very broad range of invertebrate and soil biodiversity. However, the identification and interpretation of most OTUs is not straightforward. For example, it was not possible to identify most OTUs to low-level taxonomic ranks, and many soil-derived OTUs were identified as marine organisms (e.g. Cnidaria). These problems are largely due to the inadequate representation of invertebrate biodiversity in reference sequence databases (Kwong *et al.* 2012; Kvist 2013). Furthermore, differential PCR primer binding and amplification efficiency and variable gene copy numbers between species tend to obscure the link between organism abundance/biomass and DNA sequence read abundance, making PCR-based abundance estimations problematic (Polz & Cavanaugh 1998; Porazinska *et al.* 2009; Elbrecht & Leese 2015). This is consistent with the limited correlations observed in this case between beetle specimen size/abundance and sequence abundance. PCR amplification biases tend to be predictable within species and can be mitigated with the use of degenerate primers or conserved binding sites (Krehenwinkel *et al.* 2017). This suggests that it is possible to detect shifts in relative abundance of particular organisms, and taxon-specific conversion factors would enable the determination of absolute abundances based on metabarcoding data (Krehenwinkel *et al.* 2017).

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

A lack of DNA reference data for New Zealand invertebrates currently limits the accuracy and resolution of metabarcoding. Few of the OTUs detected could be confidently identified beyond order level, due to a paucity of reference invertebrate COI sequences (Kwong *et al.* 2012; Kvist 2013). Conversely, that some OTUs could be readily identified to genus and species demonstrates the advantages of DNA-based identification, when suitable reference sequences are available. Better coverage of sequence databases is needed for the full potential of DNA metabarcoding for invertebrate biodiversity monitoring to be realised. Despite the current limitations of reference databases, and the complications posed by PCR biases, these methods can efficiently provide immense data on a very wide range of biota. With further development, DNA metabarcoding could potentially more efficiently monitor invertebrate communities, particularly for groups and habitats that are difficult to study using traditional methods. They provide data on invertebrate groups (e.g. earthworms in the soil DNA) that are neglected but have significant roles in restoring ecosystems (Snyder & Hendrix 2008). Our study applied both DNA metabarcoding and conventional methodologies to the same taxonomic groups at the same site, and in doing so, we have clarified the accuracy, consistency and limitations of DNA-based techniques. DNA metabarcoding methods offer conservation managers a practical, cost-effective technique for monitoring 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. Matt McGlone, Thomas Buckley and Anne Austin provided useful comments on the draft manuscript.

REFERENCES

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 *et al.* 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.

Clark 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.

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 *et al.* 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.

Elbrecht V & Leese F. 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence

© 2019 Australian Entomological Society

relationships with an innovative metabarcoding protocol. *PLoS One* **10**, 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**, 542–5429.
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.
Gerlach J, Samways M & Pryke J. 2013. Terrestrial invertebrates as bioindicators: an overview of available taxonomic groups. *Journal of Insect Conservation* **17**, 831–850.
Holdaway R, Wood J, Dickie I *et al*. 2017. Using DNA metabarcoding to assess New Zealand's terrestrial biodiversity. *New Zealand Journal of Ecology* **41**, 251–262.
Ji Y, Ashton L, Pedley SM *et al.* 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters* **16**, 1245–1257.
Krehenwinkel H, Wolf M, Lim JY, Rominger AJ, Simison WB & Gillespie RG. 2017. Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Scientific Reports* **7**, 17668.
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 *et al.* 2006. Soil invertebrates and ecosystem services. *European Journal of Soil Biology* **42**, S3–S15.
Leray M, Yang JY, Meyer CP *et al*. 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.
Machida RJ, Leray M, Ho S-L & Knowlton N. 2017. Metazoan mitochondrial gene sequence reference datasets for taxonomic assignment of environmental samples. *Scientific Data* **4**, 170027.
Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12.
Morinière J, Cancian de Araujo B, Lam AW *et al*. 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 *et al.* 2016. vegan: community ecology package. R package version 2.4-1.
Polz MF & Cavanaugh CM. 1998. Bias in template-to-product ratios in multitemplate PCR. *Applied and Environmental Microbiology* **64**, 3724–3730.
Porazinska DL, Giblin-Davis RM, Faller L *et al*. 2009. Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Molecular Ecolology Resources* **9**, 1439–1450.
Porter TM & Hajibabaei M. 2018. Automated high throughput animal CO1 metabarcode classification. *Scientific Reports* **8**, 4226.
R Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
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**, 629–636.
Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* **73**, 5261–5267.
Wardle DA, Bardgett RD, Klironomos JN, Setälä H, Van Der Putten WH & Wall DH. 2004. Ecological linkages between aboveground and belowground biota. *Science* **304**, 1629–1633.
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, Rohan M & Thornburrow D. 2012. Beetle community responses to grey willow (*Salix cinerea*) invasion within three New Zealand wetlands. *New Zealand Journal of Zoology* **39**, 209–227.
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* **21**, 583–598.
Wickham H. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.
Yu DW, Yinqiu J, Emerson BC *et al.* 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* **3**, 613–623.

Accepted for publication 10 December 2018.

SUPPORTING INFORMATION

Additional supporting information may/can be found online in the supporting information tab for this article.

Appendix S1 Additional details regarding study area and design.
Figure S1 Map of Mohi Bush Scenic Reserve in Hawkes Bay, New Zealand showing the locations of sampling plots.
Figure S2 Layout of sampling points within the 20 × 20 m plots.
Appendix S2 Additional details regarding DNA extraction and PCR protocols.
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.
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.
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.
Table S4 OTUs assigned to species with bootstrap confidence >90%.
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, using the Jaccard distance metric.
Figure S3 Phylum and bootstrap confidence scores for taxonomic identifications of bulk invertebrate and soil DNA OTUs using the RDP Naïve Bayesian classifier (Wang et a. 2007). Blue and red data points respectively indicate identifications from expected terrestrial invertebrate groups, and identifications from non-terrestrial invertebrates or non-invertebrate groups.

© 2019 Australian Entomological Society

Figure S4 Correlations between the number of specimens per beetle genus multiplied by mean specimen length per genus, and the square root of the number of corresponding sequence reads in each bulk invertebrate sample. Only beetle OTUs identified to Genus level that were also detected in both the conventional data are included.

Figure S5 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, using the abundance-based Bray-Curtis distance metric.

Figure S6 Non-metric MDS ordinations based on various components of the bulk invertebrate DNA and soil DNA metabarcoding datasets from forest edge and interior samples, using the abundance-based Bray-Curtis distance metric.

© 2019 Australian Entomological Society