The devil is in the detail: Metabarcoding of arthropods provides a sensitive measure of biodiversity response to forest stand composition compared with surrogate measures of biodiversity

N. Barsouma*, Catharine Bruceb, Jack Forstera, Yin-Qiu Jiic, Douglas W. Yud,

a Forest Research, Alice Holt Lodge, Farnham, Surrey, GU10 4LH, United Kingdom
b Nature Metrics CABI Site, Bakeham Lane, Egham, Surrey, TW20 9TY
c State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, 650223, China
d School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR47TJ, United Kingdom

*corresponding author: nadia.barsoum@forestry.gsi.gov.uk
ABSTRACT

Gauging trends in forest biodiversity and relating these to forest management practice and environmental change requires effective monitoring and assessment of spatio-temporal trends in forest biodiversity. Taxa- and habitat-based surrogate measures of biodiversity, or ‘biodiversity indicators’, are commonly used to convey information about the state of the biological community since they can be assessed relatively quickly and cheaply by non-experts. Direct measures of a component of biodiversity are also increasingly feasible using DNA metabarcoding; ‘Next Generation Sequencing’ has facilitated the rapid characterisation of combined multiple species samples by sequencing their DNA barcodes in parallel, simultaneously reducing the need for taxonomic expertise and the time and cost required to obtain biodiversity data across a wide range of taxonomic groups.

We investigated whether biodiversity information obtained from DNA metabarcoding of mass-trapped arthropods and from a range of taxa-based surrogate measures of biodiversity (e.g. carabid beetles, vascular plants) provide: 1) similar estimates of alpha and beta diversity and 2) provide similar forest management related conclusions. We also explored how well habitat-based surrogate measures of biodiversity (e.g. stand structure, volume of deadwood) predict observed biodiversity patterns. The study was conducted in Thetford Forest, UK within 15 forest plantation stands (5 Scots pine-oak mixtures, 4 Scots pine and 6 oak monocultures).

Our results demonstrated a high level of congruence between the metabarcoding and taxa-based surrogate measures of biodiversity. The wider range of taxonomic groups identified using a metabarcoding approach offered the potential to identify taxa sensitive to the environmental variable that was being manipulated experimentally (i.e. the composition of forest stands). Most habitat-based measures of biodiversity failed to predict species assemblage differences between stands.

Keywords: DNA metabarcoding; malaise traps; surrogate measures of biodiversity; biodiversity indicators; forest management; tree identity
1. Introduction

In recent decades there has been a growing recognition that forest management needs to balance the profitability of forest products against negative impacts on biodiversity and associated woodland ecosystem functioning and resilience (Paquette and Messier, 2010; Puettmann, 2011; Verheyen et al., 2015; Isbell et al., 2017). It is also now widely believed that with appropriate planning and management, production woodlands can play an important role in protecting and enhancing native forest biodiversity (Hartley, 2002; Quine and Humphrey, 2003; Brockerhoff et al., 2008; Gardner, 2012).

The 1992 Convention on Biological Diversity (CBD) provides a legal framework for the conservation of biodiversity and the sustainable use of its components. In the forestry sector, this stimulated the formulation of a suite of Sustainable Forest Management (SFM) principles and guidelines. These included criteria and indicators used to define SFM, but also to measure and report on progress towards the implementation of SFM (McDonald and Lane, 2004; MacDicken et al., 2015). Reflecting these catalysts of change in forest management practice, is an increasing requirement to monitor spatio-temporal trends in forest biodiversity. For example, National Forest Inventories (NFIs) now routinely include, alongside traditional measures of forest productivity, assessments designed to provide biodiversity data for national reporting against set targets to protect and enhance forest biodiversity (Chirici et al., 2012). Biodiversity data is also collected to identify woodlands of conservation interest, to detect threats (e.g. climate change, novel pests and pathogens) to forest biodiversity and to gauge the effectiveness of forest policy measures designed to enhance forest biodiversity. One such policy measure includes ‘forest diversification’ which can be achieved by fostering polycultures instead of monocultures and creating woodlands with a mixed aged structure (Puettmann, 2011).

There is common agreement among experts of the greater value of ‘actual’ compared to ‘inferred’ assessments of biodiversity (Lindenmayer and Likens, 2010; Chirici et al., 2012). Direct assessments of levels of biodiversity are, however, not straightforward. Biodiversity is broad, multidimensional, and multiscale in character making it highly challenging to monitor changes across space and time (Puulmalainen et al. 2003; Boutin et al. 2009). To census biodiversity fully, even at the smallest spatial and temporal scales, is often a prohibitively expensive and difficult task. The most common unit of taxonomic enquiry is that of the species (Hajibabaei et al., 2016) but, even at this level, biodiversity monitoring encounters numerous challenges, including: 1) the difficulty and expense of collecting representative samples of species present (e.g. trapping of rare or elusive species), 2) a shortage of taxonomic expertise to identify specimens correctly from their morphology, 3) slow processing of often very large numbers of specimens, resulting in
inevitable high related costs and 4) difficulties in identifying species due to poor quality samples, or the presence of juvenile life stages. Thus, biodiversity monitoring has tended to focus on a restricted number of species that are considered to be at risk of extinction, or species that are relatively easy to sample and that are taxonomically unambiguous and therefore easy to identify.

Alternatively, biodiversity monitoring commonly applies surrogate measures of biodiversity, or ‘biodiversity indicators’ that convey information about the wider state of the biological community and which can be assessed relatively quickly and cheaply by non-experts (Ferris and Humphrey, 1999; Noss, 1999; Coote et al., 2013). There are two categories of commonly used surrogates: taxa-based surrogates (compositional indicators) and habitat-based surrogates (structural indicators). Taxa-based surrogates refer to key taxa that are considered representative of a broader segment of biodiversity (i.e. biodiversity patterns observed for the surrogate taxon are generalizable to one or more taxa) (Sabatini et al., 2016). For example, carabid beetles (Coleoptera: Carabidae), hoverflies (Diptera: Syrphidae), spiders (Araneae), vascular plants and bryophytes are commonly cited as being potentially informative indicators of the species richness of other taxa in forest settings (Ferris and Humphrey, 1999; Cardoso et al., 2004; Pawson et al., 2011; Foord et al., 2013; Gao et al., 2015).

Habitat-based surrogates comprise aspects of the habitat that are thought to affect – and therefore predict – the richness, composition and/or diversity of one of more taxa. Examples of habitat-based surrogate measures of forest biodiversity include volumes of deadwood, levels of canopy cover and woodland stand age and structural complexity; all of these show either positive or negative correlations with species richness, depending on the taxonomic group in question (Gao et al., 2015; Tews et al., 2004). Because of the relative ease of assessing habitat-based surrogates, many of these are now included in NFIs as internationally recognised indicators of SFM and as a primary source of forest biodiversity monitoring data at the national scale (Chirici et al., 2012).

The widespread use of surrogate measures of biodiversity is, nevertheless, revealing some important limitations of these methods for forest biodiversity assessments and monitoring. Gaspar et al. (2010) cautioned that surrogate measures of biodiversity may show different strengths of correlation depending on the geographic scale of inquiry. A recent review has similarly revealed only limited evidence of the universal applicability of many commonly used surrogate measures of biodiversity in different forest ecosystems (Gao et al., 2015). This is because many have not been tested widely across different forest types and in different bioclimatic zones (Cantarello and Newton, 2008). For certain surrogate measures of biodiversity such as volume of deadwood, attempts have been made to set evidence-based threshold levels for
biodiversity gains (Humphrey and Bailey, 2012), although there is the complication that these
thresholds may need to be adjusted according to regional levels of soil fertility, the bioclimatic
zone, or depending on tree species present (Larrieu and Gonin, 2008). Furthermore, to reduce the
chances of making incorrect management decisions based on weak or ineffective surrogates that
may be biased in favour of a single taxon, several authors now recommend conducting
assessments of multiple taxonomic groups, particularly where taxonomic responses to a given
environmental variable (e.g. canopy cover) are unknown (Sabatini et al., 2015; Larrieu et al.,
2018). While this comprises a considerable sampling and sample identification effort, recent
advances in molecular ecology, and DNA metabarcoding in particular, are promising to make this
more achievable.

DNA metabarcoding is a powerful species identification method that uses ‘next generation
sequencing’ (NGS) technology to scale up the traditional DNA barcoding process. This allows the
rapid characterisation of complex samples of multiple species by sequencing their DNA barcodes
in parallel, simultaneously reducing the need for taxonomic expertise and the time and cost
required to obtain high quality biodiversity data, across a wide range of taxonomic groups, at
large spatial and temporal scales (Yu et al., 2012; Barsoum et al., 2018). Previous studies have
shown that metabarcoding arthropods generates accurate and reliable alpha and beta
biodiversity information at a fraction of the time and cost of traditional survey methods (Yu et al.,
2012; Ji et al., 2013; Morinière et al., 2016).

Here, we explore the potential to apply a metabarcoding approach to measure biodiversity
response to subtle differences in forest environmental conditions and we compare this approach
with the use of taxa- and habitat-based surrogate measures of biodiversity. Specifically, we
investigate the scope for a metabarcoding approach to provide data that can be used to: (1) detect
any fine-scale spatial and temporal variation in arthropod community composition in response to
tree species composition in plantation forest stands, (2) evaluate the biodiversity effects of
different forest management strategies; i.e. plantation monocultures compared with polycultures
and (3) identify which species or species groups of arthropods captured in malaise traps are most
sensitive to the composition of forest stands. We use a sampling method that is effective at
trapping insects from the orders Diptera and Hymenoptera (Matthews and Matthews, 1971;
Geiger et al., 2016; Morinière et al., 2016). Despite being among the most species rich groups of
arthropods, Diptera and Hymenoptera are almost always overlooked in biodiversity studies
because of the difficulty associated with sorting and identifying the inevitably large number of
specimens which tend to be characterised by small body size (Jukes and Pearce, 2003; Fraser et
al., 2008; Geiger et al., 2016).
We posed the following research questions:

(1) In forest stands of differing tree species composition, how does the information obtained from metabarcoding and from taxa-based surrogate measures of biodiversity compare? Do datasets derived from these measures of biodiversity provide similar estimates of alpha and beta diversity, thus providing similar conclusions? Taxa-based surrogate measures of biodiversity used in this study and identified based on morphology, include carabid beetles, spiders, vascular plants and bryophytes.

(2) How well do habitat-based surrogate measures of biodiversity commonly used in NFI's (e.g. stand structure, deadwood volume) predict biodiversity patterns observed by metabarcoding and taxa-based surrogate measures of biodiversity?

2. Methods

2.1. Site selection

Fifteen forest plantation stands of three stand types were selected for study: four were monocultures of Scots pine (*Pinus sylvestris* L.), six were monocultures of pedunculate oak (*Quercus robur* L.) and five were intimate mixtures of Scots pine and pedunculate oak. These were located in Thetford Forest, East Anglia in south-east England (52°30' N, 0°51' E; 10-40m a.s.l) (Thetford Forest characteristics given in Methods A1 of the Supplementary Material). The average stand size was 4.3 ha and the majority of stands were planted between 1930 and 1941 (Table 1).

Initial stand selection was based on a number of criteria: minimum stand area of 1.5ha, planting age of between 1930 and 1940, stands must have an even shape (i.e. long, thin stands with significant edge were avoided), and a stand should occur in close proximity (within the same forest management block) as selected examples of the other two stand types of interest to allow for a number of clusters of the different stand types to be sampled across the Thetford Forest region. A planting age range was selected to confine the study to a single stage of the forest harvest cycle, thus minimising the influence of stand age as a variable. Enough stands were not always found to accommodate these selection criteria, requiring two younger stands to be included (i.e. O1 and P3 planted in 1954 and 1967, respectively). The 15 stands occurred in approximately four clusters 4-12 km apart, each cluster comprising the three different plantation types.
2.2 Data collection

Biodiversity assessments comprised direct measures of biodiversity by sampling: 1) diverse taxonomic groups of flying arthropods and identifying species using metabarcoding techniques to establish the metabarcode (MBC) dataset and 2) a range of commonly used taxa-based surrogate measures of biodiversity (carabid beetles, spiders, vascular plants and bryophytes) identified based on morphology and contributing to the ‘Standard’ (STD) datasets. Indirect measures of biodiversity were also collected using habitat-based surrogate measures of biodiversity commonly used in NFI’s. These included measures of tree species composition, stand stem density and structural complexity and abundance and volume of deadwood.

2.2.1 Diverse arthropod taxa - Metabarcode (MBC) dataset

Malaise traps were used to sample sub-canopy flying arthropods. A single malaise trap was erected within a 10m radius of the centre of each stand in a space equidistant between trees, avoiding stumps, large logs and shrubs. The orientation of the malaise traps was the same in each stand; i.e. northern-most position of the trap was the main pole holding the arthropod collection vessel. Sterile collecting bottles were 2/3 filled with 100% ethanol and replaced with new ones at weekly sampling intervals for eight consecutive weeks from the 8th of August until the 4th of October 2011, giving a total of 120 (8 x 15) malaise trap samples.

2.2.2 Taxa-based surrogate measures of biodiversity - Standard (STD) datasets

Eight pitfall traps were used to sample ground-dwelling spiders and carabids in each stand (trap layout details given in Supp. Mat. Methods A2). Trap contents were collected at 7 fortnightly intervals from May to August 2011. The eight pitfall trap samples in each stand were pooled together at each sample interval. Ground-dwelling spiders and carabid beetles were identified morphologically to species level using the keys of Roberts (1993; spiders) and Luff (2007; carabids).

Vascular plants and bryophytes were surveyed in eight 2 x 2-m quadrats in each stand during the first two weeks in July 2011 (quadrat layout details given in Supp. Mat. Methods A2). The percentage cover of each terrestrial (including saxicolous and epixylic) species of vascular plant and bryophyte was estimated using the DOMIN cover-abundance scale in quadrats and the nomenclature of vascular plants and bryophytes followed Stace (2010) and Smith (2004), respectively.

2.2.3 Habitat-based surrogate measures of biodiversity

In February 2013, fourteen of the fifteen stands were surveyed to derive 16 habitat-based surrogate measures of biodiversity listed in Table 2 and described in Methods A3 (Supp. Mat.);
stand P2 could not be surveyed because it had been harvested. Definitions and assessments of stem density, deadwood and tree stumps were broadly based on those used in the UK National Forest Inventory (UK NFI, 2016).

2.3 Metabarcode protocols and data preparation

Details of sample preparation, DNA extraction, PCR and sequencing are provided in Supp. Mat. Methods A4. Methods used for the bioinformatic extraction of Operational Taxonomic Units (OTU’s) from raw sequence data are provided in Supp. Mat. Methods A5.

A total of 1123 molecular OTUs were generated, each OTU representing a distinct species. While duplicates of many of these 1123 OTUs occurred, species abundance cannot be reliably inferred from multiple identical OTUs. Quality control filtering included: 1) setting a threshold of >97% similarity match of OTU sequences, 2) the removal of single-read OTUs and 3) the removal of non-arthropods and any species with no prior record of occurrence in the UK. This reduced the number of OTUs down to 521. Of these, 67% were identifiable to species level, 8% to Genus and the remaining 25% to Order level.

Two primary metabarcode dataframes were created from the 521 OTUs that were generated from the malaise trap samples. These dataframes included a ‘binary’ dataframe and a ‘pooled’ dataframe. For the binary data frame, every OTU was scored for presence-absence in each of the 120 malaise trap samples. This dataframe was used for: 1) visualising compositional differences among samples grouped by stand type and by sample collection week (1-8) (beta diversity) and 2) for analysis of arthropod species richness between stand types (alpha diversity). In order to increase the confidence of species occurrence, single occurrence OTUs across the 120 malaise trap samples were removed from the binary dataframe.

For the pooled dataframe, where OTUs occurred in a single replicate stand, these were removed (i.e. even if an OTU was present across all eight weeks, it was excluded if it was present in only a single replicate stand). The pooled dataframe comprised species by stand data, in which the eight weekly samples were pooled within each stand. For each stand, every OTU was assigned a value between 0 and 8, representing the number of weeks in which it was detected. This index is not a direct measure of OTU abundance, but it is expected to represent each species’ contribution, over time, to a forest stand’s arthropod diversity. This dataset was used: (1) for comparisons with the STD datasets to check for consistency of between stand type trends in species richness and (2) to test for any correlations between habitat-based surrogate measures of biodiversity and beta
diversity patterns. To allow for a better comparison with the spider STD dataset, an MBC dataset was created from the pooled dataframe to include only spider OTUs ('Araneae MBC dataset').

2.4 Statistical analyses

All statistical analyses were performed using R 3.3.1 (R Core Team, 2016). The following R packages were predominantly used in the analysis: Base R package (R Core Team, 2016), Package “car” (Fox & Weisberg, 2011) for ANOVA, Package “lme4” (glmer function) (Bates et al., 2015) for Generalised linear (mixed effects) modelling (GLM/GLMM), Package “lmerTest” (Kuznetsova et al., 2014) for GLMM ANOVA, Package “lsmeans” (Lenth, 2015) for post-hoc tests least-square means, Package “mvabund” (Wang et al., 2012; Warton et al., 2012) for multivariate likelihood ratio (LR) tests, Package “multcompView” (Graves et al., 2016) for least-square means lettering and Package “vegan” (Oksanen et al., 2016) for nonmetric multidimensional scaling (NMDS) ordination.

2.4.1 Comparing species richness and community composition between stand types - MBC and STD datasets

2.4.1.1 Species richness between stand types

For the MBC dataset, total species richness per stand type was estimated using the Chao2 incidence coverage method (Chao, 1987; Colwell and Coddington, 1994), using vegan function specpool(), and compared between pairs of stand types using Welch’s t-tests. Resulting p-values were adjusted for three pairwise tests.

For the STD datasets, two metrics were used: (i) the total number of species present in each stand (TSR) (i.e. 8 quadrats/pitfall traps combined) and (ii) the mean species richness (S) per 2 x 2-m quadrat/per pitfall trap. GLMs and GLMMs with log link function and Poisson errors were used to model the effect of the explanatory variable (stand type) on the response variables (TSR, S). For mean species richness, where quadrats/pitfall traps were nested within stands, stand was used as a random effect in the mixed effects models. Since Araneae and Carabid data were collected at six intervals, collection interval was included as a factor and interaction term within the model. Where explanatory variables had a significant effect, post hoc multiple comparisons with Tukey corrections were applied.
2.4.1.2 Community composition between stand types

To visualise stand type influences on community compositions NMDS ordination of Jaccard dissimilarity matrices were created (function metaMDS() in vegan) using the MBC data. Data were displayed to show species richness differences across stand types (functions ordisurf() and ordispider() in vegan).

Multivariate LR tests were used to test for an effect of stand type on community composition across the MBC and STD data sets. In addition to testing for an overall effect of stand type, Post hoc tests were used to make pairwise comparisons between stand types, with p-values adjusted for three pairwise comparisons using Benjamini and Hochberg’s (1995) correction method (p.adjust(method=fdr) in R). Further details of the rationale and methods of applying the multivariate LR tests are given in Supp. Mat. Methods A6.

2.4.1.3 Direct comparison of MBC and STD datasets

Quantitative Jaccard distance matrices and NMDS ordinations (function metaMDS() in vegan) were created for each of the STD data sets (i.e. Araneae, Carabidae, bryophytes and vascular plants) and two MBC datasets (all arthropods and Araneae only), thereby preserving OTU frequency information. MBC and STD datasets were subsequently compared using both Procrustes and Mantel tests, each with 999 permutations, as recommended, to assess similarity between ordinations (Forcino et al., 2015).

2.4.2 Comparing habitat-based surrogate measures of biodiversity between stand types and in relation to MBC datasets

Multivariate LR tests were used to test for an effect of each of the habitat-based surrogate measures of biodiversity on community composition across the pooled arthropod MBC data, using Poisson distributions in each case. Likelihood ratio test statistics were used to determine the significance of each variable. For each variable that was significant, OTU-specific p-values and LR coefficients were used to determine the number of OTUs (by arthropod order) that showed the strongest response to the selected habitat-based surrogate measure of biodiversity.

2.4.3 Temporal variations in community composition - MBC dataset
Data were displayed using an NMDS ordination to show species richness effects across stands and time (functions ordisurf() and ordispider() in vegan). To explore time effects, data were modelled using the lmer() package in a mixed-effects model. Species richness data included all species present, including those that appeared only once within the binary data frame. Analysis of variance from the lmerTest() package (type III with Satterthwaite approximation for degrees of freedom) was used to determine significant fixed effects using a best fit model for both the MBC and Araneae MBC data. To test for differences in species associated with the first half (weeks 1-4; August) and the second half (weeks 5-8; September) of the sampling period, multivariate LR tests were conducted with binominal errors and 999 bootstrap iterations. Further details of the mixed effects model that was applied and model selection are provided in Supp. Mat. Methods A7.

3. Results

3.1 Comparing species richness and community composition between stand types - MBC and STD datasets

3.1.1 Taxonomic composition of MBC and STD datasets

MBC dataset

The 521 OTU’s making up the MBC dataset were distributed across four arthropod Classes: Arachnida, Diplopoda, Insecta and Malacostraca. Diptera were a dominant order (65% of all OTUs), followed by Coleoptera (8%), Araneae, Hemiptera, Hymenoptera (each making up 6% of all OTUs) and Lepidoptera (3%) (Table 3 and Supp. Mat. Table A1). Identification of OTU’s to species level was lowest among the Hymenoptera (52%) and Diptera (60%) and highest among better known orders such as Lepidoptera (95%), Araneae (83%) and Coleoptera (90%) which have comparatively high numbers of national recordings (NBN Atlas, 2017). Across all stands, a total of 30 spider species were identified from 10 families. Two families of spider were unique to the MBC dataset; these were orb weaver spiders (Araneidae) and mesh web weaver spiders (Dictynidae) that weave webs in vegetation. A single carabid beetle species was identified in the MBC dataset (Cyclus sp.). A number of species identified are nationally scarce or are species of declining numbers (e.g. the crab spider, Xysticus lanio; the Green-brindled Crescent moth, Allophyes oxyacanthae) and some (n = 46) from the Diptera, Hemiptera and Hymenoptera families have never previously been recorded in the Norfolk region (highlighted in Supp. Mat. Table A1). For a number of taxonomic groups (e.g. some fly and gnat families such as the Phoridae, Sciaridae, Ceratopogonidae) many species were detected that have rarely been recorded in the UK. The
MBC data also revealed the presence of a potentially important disease vector species, the biting midge *Culicoides scoticus*.

**STD datasets**

A total of 86 spider species, belonging to 17 different families, were identified in pitfall trap samples across all stands (Table Supp. Mat. Table A2). Spiders were present from eight families that did not occur in the MBC dataset. Among these were typical ground-dwelling species such as wolf (Lycosidae) and prowling (Miturgidae) spiders. A total of 37 ground-dwelling carabid species were identified from pitfall traps in all stands. Twelve of these species are frequently associated with woodlands as indicated in Supp. Mat. Table A3. A total of 67 vascular plant species and 15 bryophyte species were identified in quadrats (Supp. Mat. Tables A4 and A5, respectively).

**3.1.2 Species richness between stand types**

**MBC dataset**

No significant differences in estimated total species richness were found between oak monocultures and mixtures of Scots pine and oak, although both of these stands types had significantly higher estimated species richness than Scots pine monocultures (Figure 1). Although fewer pine monoculture stands were sampled than mixtures of Scots pine and oak, species accumulation curves indicate sufficient sampling effort for all three stand types, with the curve for Scots pine monoculture stands clearly levelling off at a lower species richness than those of the other stand types (Supp. Mat. Fig. A1).

**STD datasets**

Of the four STD datasets, only carabid and bryophyte total and mean species richness (TSR and S) showed significant differences between oak and Scots pine monocultures. There were significantly more bryophyte species, but significantly fewer carabid species in Scots pine monocultures compared with oak monocultures (Table A6). For both of these taxonomic groups, species richness in Scots pine-oak mixtures resembled the oak monocultures. In the case of spiders, a significant interaction was detected between stand type and collection interval with spider species richness in Scots pine and oak monocultures differing significantly at only one collection interval.
3.1.3 Community composition between stand types

An NMDS ordination of the MBC dataset showing arthropod samples grouped by stand type, revealed a greater similarity in the species compositions of oak monocultures and Scots pine-oak mixtures compared with Scots pine monocultures (Supp. Mat. Fig. A2). Multivariate likelihood ratio (LR) tests showed significant differences in species composition across the three stand types, with 30 OTUs associated with Scots pine-oak mixtures, 46 OTU’s associated with oak monocultures and 40 OTU’s associated with pine monocultures. These included species from a wide range of taxonomic Orders, although the majority were Diptera (Supp. Mat. Tables A1 and A7). Conifer-associated species included one potential disease vector: the biting midge Culicoides scoticus, which could be an important vector of Bluetongue virus, a serious pathogen of ruminants (Carpenter et al., 2008). The mvabund analysis showed significant differences across the three stand types for the majority of the MBC and STD data sets; pairwise comparisons of stand type are shown in Table 4. Although some of the datasets were not significant at a 0.05 level (likely due to the small sample size), there was a general trend for significant differences to be predominantly driven by pine monocultures compared with the other two stand types. The consistency across MBC and STD data sets provides evidence of consistent results across MBC and STD measures of biodiversity.

3.1.4 Direct comparison of MBC and STD datasets

Figure 2 (A-F) shows the results of the NMDS ordinations, grouped by stand type, for the MBC (Figure 2: A & B) and the STD (Figure 2: C-F) datasets. The data tend to show similar patterns, with pine monocultures being separate from the other two stand types along the primary axis. Comparison of ordinations from the Araneae pooled MBC and STD Araneae, Carabidae and vascular plant data sets indicated that the MBC and STD datasets contain similar diversity information, with significant correlation between the NMDS ordinations and Jaccard distance matrices from the MBC and STD datasets (Table 5). Comparison of ordinations from the total pooled MBC dataset and the bryophyte STD dataset and comparison of the Araneae pooled MBC dataset and the STD Araneae dataset indicated that the MBC and STD datasets may contain similar diversity information, with significant correlation between the NMDS ordinations but not the Jaccard distance matrices from the MBC and STD datasets; this latter lack of correlation may be related to the limited number of spiders identified in the MBC dataset.
3.2 Comparing habitat-based surrogate measures of biodiversity between stand types and in relation to MBC datasets

The mvabund analysis showed significant differences across only one of the surrogate variables: percentage of pine cover (community ~ perc_pine (Poisson errors), Dev(1,13) = 1,480, p = 0.02). OTU-specific p-values and LR coefficients were used to determine the number of OTUs (by arthropod order) that showed the strongest response to percentage pine (Table 6), with Diptera and Araneae being the predominant orders showing a response. Figure 3 shows a heat map plot of the arthropod MBC data arranged by stand type and % of pine within each stand, showing how different taxa are driving community differences between stand types. Sites P2 and P4 feature particularly distinct arthropod communities. These are pure pine monocultures that lack broadleaf trees even in the understory.

3.3 Temporal variations in community composition - MBC dataset

Analysis of variance applied to the mixed effect model indicated no significant effects of stand type or the interaction between stand type and time (days) (Figure 4). When the same best fit model was applied to Araneae only MBC data, these data would not converge even with the increased number of dimensions. Analysis of the second NMDS dimension by week as a factor*stand type showed significant main effects with no interaction, where week as a response was non-linear (Figure A3). Splitting the data into two halves (weeks 1 to 4 and weeks 5 to 8) identified 53 OTUs as being strongly associated with the first half of the trapping period and 54 with the second half. The majority of species driving the temporal effect were dipterans, along with several hymenopteran species (Table A8). Associations are consistent with the species biology. For example, the moth species Tischeria ekebladella (associated with weeks 1-4) typically flies in the summer, entering a larval stage from September. Similarly, the ant species Myrmica ruginodis was detected in several stands during the first three trapping weeks, after which it was never detected; this is consistent with mating flights for this species which occur in July and August.

4. Discussion

4.1. MBC and STD datasets of multiple taxonomic groups show similar alpha and beta diversity trends across different stand types with comparable forest management implications
The MBC and STD datasets both showed a distinctiveness in the composition of communities sampled in Scots pine monocultures compared with oak monocultures for all taxonomic groups assessed. In Scots pine-oak mixed stands, MBC and STD datasets also showed the same tendency for communities to occupy an “intermediate” position in ordinations, with communities partially comprised of component species present in either Scots pine or oak monocultures. These results are in line with a growing number of studies demonstrating the effectiveness of DNA metabarcoding as a method of collecting reliable biodiversity information that can be used to inform management practice and policy (Ji et al., 2013; Deiner et al., 2017; Elbrecht et al., 2017).

In this study, the data provides evidence backing current UK forestry policy that advocates a diversification in the composition of forest stands and woodlands for biodiversity gains (FC, 2017). Thetford Forest is dominated by pine and these results suggest that the inclusion of oak stands as part of the wider mosaic of woodland stands would improve overall levels of alpha and beta diversity. A notable result is the limited ordination space occupied by Scots pine-oak mixtures compared with oak and Scots pine monocultures combined, with mixed stands particularly failing to cover the space occupied by pine monocultures (Figure 3). This suggests that in oak and Scots pine plantations, improved regional species diversity (for the taxonomic groups considered here) can be achieved by creating a mosaic of pure-oak and pure-pine crops rather than planting intimate mixtures of Scots pine and oak; this is because Scots pine-oak mixtures would incur the loss of pine specialists.

In the Thetford Forest context, Scots pine and oak were clearly favoured by different taxonomic groups; i.e. spiders and bryophytes showed significantly higher species richness in Scots pine monocultures compared with oak monocultures, while carabid beetles showed higher species richness in oak monocultures. There is a need, however, to be cautious about how transferable these taxa-specific responses are in different spatial and temporal contexts. For example, we did not find significant differences in spider species richness between stand types across all sampling intervals. Identical responses have also not been found for many of these taxonomic groups (i.e. vascular plants, spiders, carabids) in other regions of study when comparing these same stand types (Taboda et al., 2010; Barsoum et al., 2016). This inconsistency in taxa-based surrogate measures of biodiversity in different climatic and biogeographical contexts has been reported elsewhere and points to the limitations of focussing biodiversity monitoring and assessment on a single taxa-based surrogate measure of biodiversity, but also over a restricted sampling interval (Kirkman et al., 2012; Sabatini et al., 2016).
4.2. The MBC dataset is more taxonomically comprehensive than STD datasets, allowing for a greater number and range of species associations to be identified by stand type than individual taxa-based surrogate measures of biodiversity.

The use of malaise traps and subsequent species identification by metabarcoding allowed for a comparatively large number of species to be sampled across numerous taxonomic groups (particularly among the hyper-diverse Diptera). This improved the chances of identifying whole taxonomic groups that show a particular sensitivity to tree identity, but also individual arthropod species with particular stand type associations; i.e. a total of 116 arthropod species from the MBC dataset had particular stand type associations. For example, high proportions of the dark-winged fungus gnats (Sciariidae) sampled were found to have a significant association to a single stand type. This highlights the scope for the metabarcoding approach to identify taxa-based indicators in forests that demonstrate a particular sensitivity to a given environmental characteristic (e.g. in this case, tree species). It follows that this opens up the possibility of developing and applying metabarcoding as a comparatively rapid and inexpensive tool for routine monitoring (Morinère et al., 2016) in a similar way to current achievements in freshwater ecosystems. Freshwater ecologists are striving and making good progress in the use of DNA metabarcoding of macroinvertebrates to monitor instream water quality (Elbrecht et al., 2017). While species level identification may not be possible for all arthropod specimens sampled due to biases introduced by primers used and reference barcode library limitations the range and number of arthropod species that can be identified using a metabarcoding approach are nevertheless highly informative and are increasing all the time. Molecular methods have already advanced significantly since we completed the molecular work on our study and yet even with the lower resolution we used compared to what is currently achievable with greater sequencing depth, we were able to detect species: 1) of conservation interest (e.g. Green-brindled Crescent moth, *A. oxyacanthae*), 2) that may pose a biosecurity risk (e.g. the biting midge *C. scoticus* as a potential pathogen vector) and 3) that have not previously been recorded in the region of study. Key to building a monitoring platform using metabarcoding, however, will be the need to standardise sampling and analytical methods for directly transferable and comparable biodiversity estimates (Cristescu, 2014). This is especially vital where it is envisioned that DNA metabarcoding is applied as a monitoring tool for use within legal and regulatory frameworks (Leese et al., 2018).

The careful selection of primers is an additional requirement. Since completing our study, Morinère et al. (2016) have published a study comparing the efficiency of different primers using arthropod samples captured in a malaise trap. Primers used in our study were among those tested by Morinère et al. (2016) who found greater efficiency of amplicons using the dgHCO primer (Leray et al., 2013) than the two primers used in our study; i.e. LCO1490 and HCO2198.
(Folmer et al., 1994). This may go some way to explain the surprisingly low proportions of Hymenoptera detected in our study and another malaise trap study that also used Folmer’s primers (Yu et al., 2012).

4.2. Most habitat-based surrogate measures of biodiversity tested did not predict significant differences in species assemblages between stands

While some difference in structural complexity and deadwood volume were expected between the different stand types based on the differing characteristics of the tree species (Mason and Connolly, 2014; Shorohova and Kapitsa, 2014; Herrmann et al., 2015, Pretzsch, 2017), these differences were not captured by the variables measured in this study. The range of UK-NFI habitat-based surrogate measures of biodiversity that were assessed revealed a consistency in the measured habitat conditions across the different stands and stand types. Stem density, stand structural complexity, levels of deadwood and the number of canopy and sub-canopy tree species were comparable across the stands and thus, were not useful predictors of significant species and compositional differences observed in the MBC and STD datasets between the different stand types. Only one variable was found to reflect the compositional differences in arthropod communities found in the different stand types based on the MBC dataset; that was the percentage of conifer (i.e. Scots pine) as a proportion of all trees present in the stand. These results suggest that a reliance on the habitat-based surrogate measures of biodiversity applied here would have led to incorrect assumptions being made about underlying patterns of biodiversity (e.g. significant differences in patterns of species richness between the different forest stand types might have been overlooked).

4.3. Metabarcoding captures fine-scale temporal variations in the composition of arthropod communities

Arthropod sampling can very quickly generate extremely large, unwieldy numbers of specimens, particularly less targeted sampling techniques such as malaise traps. This greatly restricts the number of taxa and repeat samples than can be processed where species identification is based on morphology alone (Humphrey et al., 2003; Morinie, et al., 2016). Identification of species using the metabarcoding approach made it possible for a high intensity and frequency of arthropod assemblages to be processed. This provided insight into the very rapid changes in composition of arthropod communities over an eight week period within each stand. Our results showed similar rates of species assemblage change across stands and clear species associations
with different sampling periods indicating evident compositional shifts through time. These findings underline the importance of controlling for temporal effects in sampling using malaise traps, and particularly for certain taxonomic groups such as parasitoid wasps; the species composition of samples collected just a couple of weeks apart can differ greatly (Fraser et al., 2008; Geiger et al., 2016). Our findings additionally highlight the potential to relate finely-grained temporal shifts in arthropod communities to fluctuating environmental variables in order to explain the root causes of important shifts in the composition of arthropod communities. This is particularly relevant when considering significant reported global declines in the abundance of certain insect groups, including moths, butterflies, bees, spiders and carabid beetles (Hallmann et al., 2017; Leather, 2018). The causal agents of many of these declines are not yet clear, although environmental variables with a negative influence could include levels of air pollution and pesticide use associated with land use intensification, and/or important variations in the seasonality and range of ambient temperatures associated with global warming (Brandon-Mong et al., 2018).

Acknowledgements

This work has been jointly sponsored by the Forestry Commission and the European Union (European Regional Development Fund ERDF) within the framework of the Forestry Commission Climate Change Adaptation and Biodiversity Programmes and the European INTERREG IV A 2 Mers Seas Zeeën Cross-border Cooperation Programme 2007–2013 (Project 090316016-FR MULTIFOR: Management of Multi-Functional Forests). This work also benefitted from a NERC PhD studentship grant awarded to Catherine Bruce at the University of East Anglia (2010–2013). We gratefully acknowledge the contributions of Chen-Xue Yang, Amy Eycott, Alex Robinson, Lauren Fuller and Forest Research’s Technical Support Unit in the sample processing stages of the work.

References


mixed tree species plantations compare with monocultures? Forest Ecology and Management, 382, 244-256.


transforming how we survey animal and plant communities. Molecular Ecology, 26(21), 5872-5895. doi.org/10.1111/mec.14350


Mason, W.L., Connolly, T., 2014. Mixtures with spruce species can be more productive than monocultures: evidence from the Gisburn experiment in Britain. Forestry, 87, 209-217.


global network of tree diversity experiments to sustainable forest plantations. Ambio, 45, 29-41 - doi:10.1007/s13280-015-0685-1


List of Tables:

Table 1. Summary characteristics of 15 study stands in the Thetford Forest region.

Table 2: Names and descriptions of habitat-based surrogate measures of biodiversity included in study.

Table 3: Taxonomic composition of MBC dataset

Table 4. Results of Multivariate LR tests applied to MBC and STD datasets, comparing each stand type separately. P-values (p) are adjusted for three tests using Benjamini and Hochberg’s (1995) correction. Significant associations with stand type are shown in bold italics.

Table 5: Comparison of MBC and STD datasets; i.e. level of correlation between NMDS ordinations and Jaccard distances matrices.

Table 6. Number of OTUs in each taxonomic group that are significantly associated with percentage of pine in a stand.

List of Figures:

Figure 1: Estimated extrapolated species richness (alpha diversity) of all arthropods combined (MBC dataset) in Scots pine oak mixed stands, and in oak and Scots pine monocultures calculated using the Chao equation. Error bars indicate standard errors.

Figure 2. Non-metric multidimensional scaling (NMDS) ordinations (A-F) of MBC datasets (all arthropods, Araneae only) and STD datasets (spiders, carabids, vascular plants, bryophytes) showing samples grouped by stand type. Surface plot shows species richness.

Figure 3. Pooled total MBC data as a heat map plot. Stands are arranged by percentage of pine present (low to high) on the x-axis. Occurrence of different OTUs are represented by coloured lines on the y-axis.

Figure 4. NMDS ordination showing MBC samples (all arthropods) grouped by week. Surface plot shows species richness.

Appendix A: Supplementary material
Additional Supplementary material describing Methods associated with this article are listed below.

**Methods A1**: Site selection (2.1)

**Methods A2**: Taxa-based surrogate measures of biodiversity - Standard (STD) datasets (2.2.2)

**Methods A3**: Habitat-based surrogate measures of biodiversity (2.2.3)

**Methods A4**: Sample preparation, DNA extraction, PCR and sequencing (2.3.1)

**Methods A5**: Bioinformatic extraction of operational taxonomic units (OTU’s) from raw sequence data (2.3.2)

**Methods A6**: Community composition between stand types (2.4.1.2)

**Methods A7**: Temporal variations in community composition – MBC dataset (2.4.3)

Additional Supplementary figures associated with this article are listed below.

**Figure A1**: Species accumulation curves for mixtures of Scots pine and oak (red), oak monocultures (green) and pine monocultures (blue), estimated using specaccum() function in vegan() package in R. Method = “exact” (finds the expected (mean) species richness), permutations = 9999.

**Figure A2**: NMDS ordination showing MBC arthropod samples grouped by stand type. Surface plot shows species richness.

**Figure A3**: Boxplot of second NMDS dimension by week indicating a non-linear response (flattening from ~week 5 onwards).

Additional Supplementary tables associated with this article are listed below.

**Table A1**: List of species/OTUs in the MBC dataset. Occurrence is indicated by stand type. Also indicated are species/OTUs with significant tree species associations. Species not previously recorded in the region of study (Norfolk) are highlighted with an asterisk.

**Table A2**: List of spider species present in each stand type. All pitfall trap data for each given stand type combined.

**Table A3**: List of carabid species present in each stand type. All pitfall trap data for each given stand type combined.
Table A4: List of vascular plant species present in each stand type. All 2m x 2m quadrat data for each given stand type combined.

Table A5: List of bryophyte species present in each stand type. All 2m x 2m quadrat data for each given stand type combined.

Table A6: Mean total species richness (TSR) and mean species richness (S) of Araneae, Carabidae, vascular plants and bryophytes (STD datasets) in Scots pine-oak mixed (SP/OK) and monoculture (SP, OK) stands. Standard error is given in brackets. Different lower case letters indicate a significant difference (p < 0.05–0.001) between stand types.

Table A7: Number of OTUs in each taxonomic Order that are significantly associated with each stand type. Based on three separate multivariate LR tests in mvabund with binomial errors, malaise.trap resampling and 999 bootstrap iterations. Each analysis tested one stand type against the other two (pooled).

Table A8: Number of OTUs in each taxonomic group that are significantly associated with the first half (weeks 1-4; August) and second half (weeks 5-8; September) of the sampling period. Based on LR tests in mvabund with binomial errors, pit.trap resampling and 999 bootstrap iterations.
Figure 1: Estimated extrapolated species richness (alpha diversity) of all arthropods combined (MBC dataset) in Scots pine oak mixed stands, and in oak and Scots pine monocultures calculated using the Chao equation. Error bars indicate standard errors.
Figure 2. Non-metric multidimensional scaling (NMDS) ordinations (A-F) of MBC datasets (all arthropods, Araneae only) and STD datasets (spiders, carabids, vascular plants, bryophytes) showing samples grouped by stand type. Surface plot shows species richness.

A - MBC, All arthropods, malaise traps
B - MBC, Araneae, malaise traps

C - STD, Carabidae, pitfall traps
D - STD, Araneae, pitfall traps

E - STD, bryophytes, quadrats
F - STD, vascular plants, quadrats
**Figure 3.** Pooled total MBC data as a heat map plot. Stands are arranged by percentage of pine present (low to high) on the x-axis. Occurrence of different OTUs are represented by coloured lines on the y-axis.
**Figure 4.** NMDS ordination showing MBC samples (all arthropods) grouped by week. Surface plot shows species richness.
Table 1. Summary characteristics of 15 study stands in the Thetford Forest region.

<table>
<thead>
<tr>
<th>Site code</th>
<th>Site history+ Landcover 1905 -1910</th>
<th>Current stand type* (% Pine)</th>
<th>Planting year</th>
<th>Stand Area (ha)</th>
<th>Altitude (m a.s.l.)</th>
<th>Soil type</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>C/B mix</td>
<td>OK/SP (20)</td>
<td>1941</td>
<td>4.9</td>
<td>25</td>
<td>Brown Earth</td>
</tr>
<tr>
<td>M2</td>
<td>C/B mix</td>
<td>OK/SP (74)</td>
<td>1932</td>
<td>3.4</td>
<td>15</td>
<td>Brown Earth</td>
</tr>
<tr>
<td>M4</td>
<td>Bare</td>
<td>OK/SP (40)</td>
<td>1934</td>
<td>4.5</td>
<td>30</td>
<td>Brown Earth</td>
</tr>
<tr>
<td>M5</td>
<td>Bare</td>
<td>OK/SP (45)</td>
<td>1932</td>
<td>5.2</td>
<td>40</td>
<td>Brown Earth</td>
</tr>
<tr>
<td>M6</td>
<td>Bare</td>
<td>OK/SP (24)</td>
<td>1935</td>
<td>5.2</td>
<td>40</td>
<td>Ground Water Gley</td>
</tr>
<tr>
<td>O1</td>
<td>Bare</td>
<td>OK (0)</td>
<td>1954</td>
<td>4.7</td>
<td>10</td>
<td>Loamy Texture</td>
</tr>
<tr>
<td>O2</td>
<td>Bare</td>
<td>OK (0)</td>
<td>1934</td>
<td>4.9</td>
<td>25</td>
<td>Calcareous Brown Earth</td>
</tr>
<tr>
<td>O3</td>
<td>Bare</td>
<td>OK (0)</td>
<td>1934</td>
<td>2.4</td>
<td>35</td>
<td>Brown Earth</td>
</tr>
<tr>
<td>O4</td>
<td>Bare</td>
<td>OK (0)</td>
<td>1933</td>
<td>2.9</td>
<td>20</td>
<td>Brown Earth</td>
</tr>
<tr>
<td>O5</td>
<td>Bare</td>
<td>OK (0)</td>
<td>1932</td>
<td>6.8</td>
<td>40</td>
<td>Brown Earth</td>
</tr>
<tr>
<td>O6</td>
<td>C/B mix</td>
<td>OK (3)</td>
<td>1934</td>
<td>5.2</td>
<td>20</td>
<td>Calcareous Brown Earth</td>
</tr>
<tr>
<td>P1</td>
<td>Bare</td>
<td>SP (100)</td>
<td>1930</td>
<td>1.7</td>
<td>30</td>
<td>Brown Earth</td>
</tr>
<tr>
<td>P2</td>
<td>Bare</td>
<td>SP (100)</td>
<td>1941</td>
<td>1.6</td>
<td>30</td>
<td>Typical Podzol</td>
</tr>
<tr>
<td>P3</td>
<td>C/B mix</td>
<td>SP (100)</td>
<td>1967</td>
<td>3.6</td>
<td>30</td>
<td>Brown Earth</td>
</tr>
<tr>
<td>P4</td>
<td>Bare</td>
<td>SP (100)</td>
<td>1937</td>
<td>7.1</td>
<td>35</td>
<td>Calcareous Brown Earth</td>
</tr>
</tbody>
</table>

+ Land cover classes include conifer woodland (C), broadleaf woodland (B), conifer and broadleaf mixed woodland (C/B mix) and non-wooded areas (Bare) that could in some cases be areas of heathland.

*Three stand types: OK/SP = mixture, OK = oak monoculture, SP=Scots pine monoculture.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree species</td>
<td>Number of tree species with at least one measurable stem</td>
</tr>
<tr>
<td>%Pine</td>
<td>Percentage of measurable stems (crop and non-crop; live and dead) that are Scots pine. A measure of the broadleaf/conifer ratio</td>
</tr>
<tr>
<td>Stem density</td>
<td>Number of measurable stems (live and dead) in 900m$^2$ block</td>
</tr>
<tr>
<td>Crop density</td>
<td>Number of crop stems (i.e. Scots pine and/or oak) in 900m$^2$ block</td>
</tr>
<tr>
<td>Non-crop density</td>
<td>Number of non-crop stems in 900m$^2$ block; i.e. non-canopy Scot spine and/or oak and other tree species present</td>
</tr>
<tr>
<td>SCI</td>
<td>Structural complexity index (Zenner and Hibbs, 2000)</td>
</tr>
<tr>
<td>ESCI 1</td>
<td>Enhanced SCI, modification step 1 (ESCI'). Incorporates triangle orientations (Beckschäfer et al., 2013)</td>
</tr>
<tr>
<td>ESCI 2</td>
<td>Enhanced SCI, modification step 2 (ESCI). Incorporates triangle orientations and stem density (Beckschäfer et al., 2013)</td>
</tr>
<tr>
<td>Simpson count</td>
<td>Simpson's diversity index D for trees, based on count of measurable stems</td>
</tr>
<tr>
<td>Simpson area</td>
<td>Simpson's diversity index D for trees, based on cross-sectional area of measurable stems</td>
</tr>
<tr>
<td>Deadwood area</td>
<td>Total cross-sectional area of lying deadwood stems intersecting transect line</td>
</tr>
<tr>
<td>Deadwood count</td>
<td>Number of lying deadwood pieces intersecting transect lines</td>
</tr>
<tr>
<td>Stump area</td>
<td>Total cross-sectional area of stumps in circular plots based on stump height and diameter</td>
</tr>
<tr>
<td>Stump count</td>
<td>Total number of stumps in circular plots</td>
</tr>
<tr>
<td>DS area</td>
<td>Deadwood area + Stump area</td>
</tr>
<tr>
<td>DS count</td>
<td>Deadwood count + Stump count</td>
</tr>
</tbody>
</table>
Table 3: Taxonomic composition of MBC dataset

<table>
<thead>
<tr>
<th>Class</th>
<th>Order</th>
<th>Number of species/ OTUs</th>
<th>Percentage of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arachnida</td>
<td>Araneae</td>
<td>30</td>
<td>5.7</td>
</tr>
<tr>
<td></td>
<td>Opiliones</td>
<td>5</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Sarcoptiformes</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>Diplopoda</td>
<td>Julida</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>Insecta</td>
<td>Coleoptera</td>
<td>39</td>
<td>7.5</td>
</tr>
<tr>
<td></td>
<td>Dermaptera</td>
<td>2</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>Diptera</td>
<td>338</td>
<td>64.8</td>
</tr>
<tr>
<td></td>
<td>Hemiptera</td>
<td>29</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>Hymenoptera</td>
<td>31</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td>Lepidoptera</td>
<td>18</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>Mecoptera</td>
<td>3</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>Neuroptera</td>
<td>6</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>Orthoptera</td>
<td>5</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Plecoptera</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>Psocodea</td>
<td>8</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>Psocoptera</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>Trichoptera</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>Malacostraca</td>
<td>Isopoda</td>
<td>2</td>
<td>0.4</td>
</tr>
</tbody>
</table>
Table 4. Results of Multivariate LR tests applied to MBC and STD data sets, comparing each stand type separately. P-values ($p$) are adjusted for three tests using Benjamini and Hochberg’s (1995) correction. Significant associations with stand type are shown in bold italics.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Overall $p$</th>
<th>Oak $p$</th>
<th>Pine $p$</th>
<th>Mix $p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled all arthropods MBC</td>
<td>0.05</td>
<td>0.23</td>
<td>0.09</td>
<td>0.40</td>
</tr>
<tr>
<td>Araneae MBC</td>
<td>0.03</td>
<td>0.05</td>
<td>0.05</td>
<td>0.63</td>
</tr>
<tr>
<td>Pooled pitfall STD</td>
<td>0.01</td>
<td>0.09</td>
<td>0.02</td>
<td>0.37</td>
</tr>
<tr>
<td>Araneae pitfall STD</td>
<td>0.01</td>
<td>0.05</td>
<td>0.04</td>
<td>0.27</td>
</tr>
<tr>
<td>Carabidae pitfall STD</td>
<td>0.02</td>
<td>0.46</td>
<td>0.03</td>
<td>0.47</td>
</tr>
<tr>
<td>Bryophyte STD</td>
<td>0.02</td>
<td>0.13</td>
<td>0.02</td>
<td>0.50</td>
</tr>
<tr>
<td>Vascular plants STD</td>
<td>0.12</td>
<td>0.34</td>
<td>0.27</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Table 5: Comparison of MBC and STD datasets; i.e. level of correlation between NMDS ordinations and Jaccard distances matrices.

<table>
<thead>
<tr>
<th>MBC dataset</th>
<th>STD dataset</th>
<th>Procrustes test correlation</th>
<th>Mantel test r</th>
</tr>
</thead>
<tbody>
<tr>
<td>All arthropods</td>
<td>Araneae</td>
<td>0.68**</td>
<td>0.31**</td>
</tr>
<tr>
<td>Araneae</td>
<td>Araneae</td>
<td>0.65**</td>
<td>0.14+</td>
</tr>
<tr>
<td>All arthropods</td>
<td>Carabidae</td>
<td>0.58**</td>
<td>0.27*</td>
</tr>
<tr>
<td>All arthropods</td>
<td>Bryophytes</td>
<td>0.53*</td>
<td>0.18+</td>
</tr>
<tr>
<td>All arthropods</td>
<td>Vascular plants</td>
<td>0.56**</td>
<td>0.30*</td>
</tr>
</tbody>
</table>

Significance level indicated by +<0.1, *<0.05, **<0.01.
**Table 6.** Number of OTUs in each taxonomic group that are significantly associated with percentage of pine in a stand.

<table>
<thead>
<tr>
<th>Order</th>
<th>Number of OTU’s associated with % pine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Araneae</td>
<td>9</td>
</tr>
<tr>
<td>Opiliones</td>
<td>2</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>4</td>
</tr>
<tr>
<td>Diptera</td>
<td>39</td>
</tr>
<tr>
<td>Hemiptera</td>
<td>2</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>4</td>
</tr>
<tr>
<td>Lepidoptera</td>
<td>4</td>
</tr>
<tr>
<td>Neuroptera</td>
<td>2</td>
</tr>
<tr>
<td>Orthoptera</td>
<td>1</td>
</tr>
<tr>
<td>Psocodea</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>69</strong></td>
</tr>
</tbody>
</table>