ECOLOGY LETTERS

Ecology Letters, (2013) **16**: 1245–1257 doi: 10.1111/ele.12162

IDEA AND PERSPECTIVE

Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding

Yinqiu Ji,[1‡] Louise Ashton,[2‡] Scott M. Pedley,[3‡] David P. Edwards,[4,5,6‡,†] Yong Tang,[7] Akihiro Nakamura,[2,8] Roger Kitching,[2] Paul M. Dolman,[3] Paul Woodcock,[9] Felicity A. Edwards,[9] Trond H. Larsen,[10] Wayne W. Hsu,[11] Suzan Benedick,[12] Keith C. Hamer,[9] David S. Wilcove,[4,6] Catharine Bruce,[13] Xiaoyang Wang,[1] Taal Levi,[14,15] Martin Lott,[16] Brent C. Emerson[17] and Douglas W. Yu[1,13]*

Abstract

To manage and conserve biodiversity, one must know what is being lost, where, and why, as well as which remedies are likely to be most effective. Metabarcoding technology can characterise the species compositions of mass samples of eukaryotes or of environmental DNA. Here, we validate metabarcoding by testing it against three high-quality standard data sets that were collected in Malaysia (tropical), China (subtropical) and the United Kingdom (temperate) and that comprised 55,813 arthropod and bird specimens identified to species level with the expenditure of 2,505 person-hours of taxonomic expertise. The metabarcode and standard data sets exhibit statistically correlated alpha- and beta-diversities, and the two data sets produce similar policy conclusions for two conservation applications: restoration ecology and systematic conservation planning. Compared with standard biodiversity data sets, metabarcoded samples are taxonomically more comprehensive, many times quicker to produce, less reliant on taxonomic expertise and auditable by third parties, which is essential for dispute resolution.

Keywords

Biodiversity, climate change, DNA barcoding, heathland, restoration ecology, surveillance monitoring, systematic conservation planning, targeted monitoring, tropical forest.

Ecology Letters (2013) **16**: 1245–1257

Quantity has a quality all of its own.

Joseph Stalin

INTRODUCTION

Many of the challenges of biodiversity conservation can be thought of as problems of management, and in management, it is a truism that you only get what you measure. Efforts to design efficient biodiversity indicators that are useful for management (e.g. Pereira *et al.* 2013), and arguments over the allocation of effort to monitoring versus action (e.g. Knight *et al.* 2010; Stuart *et al.* 2010), are therefore active (and contentious) research themes in conservation science.

As just one example of the usefulness of indicators, bushmeat hunting is a well-known biodiversity threat. In Amazonian rainforest, it is not feasible to monitor hunter effort, but because human population densities are low, it is possible to create long-term hunting refuges for game species by using infrastructure investments, like potable water systems, to encourage existing human settlements to grow and to discourage the creation of new settlements. Unlike hunter behaviour, settlements are visible and verifiable indicators that happen to strongly predict the distribution of hunting effort. Management can therefore monitor settlements as a proxy for hunting pressure and use infrastructure investment as a self-enforcing payment for foregone hunting, because new settlements forgo benefits (Levi *et al.* 2009; Yu 2010).

However, there is a substantial literature that has critiqued the use of indicators and umbrella species for biodiversity monitoring and environmental management (e.g. Andelman & Fagan 2000; Cushman *et al.* 2010; Stuart *et al.* 2010; Lindenmayer & Likens 2011; Newton 2011; Dolman *et al.* 2012; Nicholson *et al.* 2012 and

[1]*State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, 650223, China*
[2]*Environmental Futures Centre and Griffith School of Environment, Griffith University, Nathan, Queensland, 4111, Australia*
[3]*School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR47TJ, UK*
[4]*Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, New Jersey, 08544, USA*
[5]*Centre for Tropical Environmental and Sustainability Science (TESS), School of Marine and Tropical Biology, James Cook University, Cairns, Queensland, 4878, Australia*
[6]*Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey, 08544, USA*
[7]*Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Meng La, Yunnan, 666303, China*
[8]*Queensland Museum, South Brisbane, Queensland, 4101, Australia*
[9]*Institute of Integrative and Comparative Biology and Faculty of Biological Sciences, University of Leeds, Leeds, West Yorkshire, LS29JT, UK*
[10]*The Betty and Gordon Moore Center for Ecosystem Science and Economics, Conservation International, Arlington, Virginia, 22202, USA*
[11]*Department of Ecology, Evolution, and Environmental Biology, Columbia University, New York City, New York, 10027, USA*
[12]*School of Sustainable Agriculture, Universiti Malaysia Sabah, Sandakan, Sabah, 9000, Malaysia*
[13]*School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR47TJ, UK*
[14]*Cary Institute of Ecosystem Studies, Millbrook, New York, 12545, USA*
[15]*Department of Biology, University of Florida, Gainesville, Florida, 32611, USA*
[16]*School of Computing Sciences, University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR47TJ, UK*
[17]*Island Ecology and Evolution Research Group, IPNA-CSIC, Tenerife, Canary Islands, 38206, Spain*
**Correspondence. E-mail: dougwyu@gmail.com*
†*Present address: Department of Animal and Plant Sciences, University of Sheffield, Sheffield, S10 2TN UK.*
‡*These authors contributed equally to this study.*

included references). Wiens *et al.* (2009) warn that remote sensing 'is not a panacea for the challenges of conducting ecological monitoring….' Gardner *et al.* (2012) highlight the need for high-quality, on-the-ground data in validating remote-sensing indicators for Reducing Emissions from Deforestation and Forest Degradation (REDD) projects that are aimed at both carbon and biodiversity protection – particularly where anthropogenic impacts are more subtle than habitat conversion.

Thus, a complementary approach to reliance on indicators is to devise technologies to monitor policy targets *directly*. An illustrative example is airborne LiDAR sensing (Asner *et al.* 2010), which directly provides large-scale, high-resolution forest-carbon estimates that can be used to properly target payments for REDD projects. For direct biodiversity measurement, the leading technological candidate is metabarcoding (Baird & Hajibabaei 2012; Bik *et al.* 2012; Taberlet *et al.* 2012; Yu *et al.* 2012), which applies microbial metagenetic technology to eukaryotes (Box 1). Amplicons of species-discriminating 'barcode' genes from soil, water, air or collections of organisms provide presence/absence data for plants, invertebrates and vertebrates (Fonseca *et al.* 2010; Hajibabaei *et al.* 2011; Thomsen *et al.* 2011; Hiiesalu *et al.* 2012; Yoccoz *et al.* 2012; Yu *et al.* 2012) and can recover ecological information in the form of alpha- and beta-diversity estimates (Fonseca *et al.* 2010; Hiiesalu *et al.* 2012; Yoccoz *et al.* 2012; Yu *et al.* 2012). 'Meta' refers to the 'collective' study of all barcode genes present in a sample (Box 1).

Importantly, such collections are auditable, because sites can be sampled by independent parties, or samples can be split, and analysed by certified entities following a standardised protocol. They can also be verified (at extra cost) by fieldwork to confirm the presence or absence of particular species. Metabarcode data sets are also taxonomically more comprehensive, many times quicker to produce, and less reliant on taxonomic expertise (Baird & Hajibabaei 2012; Bik *et al.* 2012; Taberlet *et al.* 2012; Yu *et al.* 2012).

However, despite these advantages, it is not yet the case that metabarcode data sets can be treated as *reliable* sources of biodiversity information for policymaking. All previous validations of metabarcoding, including our own, have been tested against laboratory-assembled samples of known composition (e.g. Porazinska *et al.* 2009; Hajibabaei *et al.* 2012; Hiiesalu *et al.* 2012; Yu *et al.* 2012; Kermarrec *et al.* 2013; Zhan *et al.* 2013; Zhou *et al.* 2013) or have

Box 1 What is metabarcoding?

Metabarcoding is a rapid method of biodiversity assessment that combines two technologies: DNA taxonomy and high-throughput DNA sequencing.

Short sequences of DNA are widely used to differentiate and assign taxonomies to specimens of animals, plants, and fungi and other microbes. For animals, the most commonly used sequence is a 658-base-pair portion of the mitochondrial cytochrome oxidase subunit I gene, or COI, which is known as a 'DNA barcode.' Other barcode genes are used for fungi and plants. A simple introduction to barcodes is available at www.barcodeoflife.org (accessed 17 May 2013).

So-called 'barcoding campaigns' are managed through the Barcode of Life Database (BOLD) (Ratnasingham & Hebert 2007). Official barcode sequences are tied to a curated specimen deposited in a museum and meet certain metadata standards, the intent being to provide auditable taxonomies. Barcode sequences are indicated by a BARCODE tag and are deposited permanently in the International Nucleotide Sequence Database Collaboration (INSDC), which comprises GenBank in the US, the DNA Data Bank of Japan (DDBJ) and the European Nucleotide Archive (ENA) (Table 4). A variety of taxonomically informative genes other than barcodes are curated in specialised databases (Table 4). All these data sources, plus the many sequences generated from general scientific research and uploaded directly to the INSDC databases, can be used for taxonomic assignment in metabarcoding studies.

Genes used for taxonomy should have at least two important properties. First, they should mutate at just the right rate so that sequences in different species differ by at least a few percentage points worth of base pairs (typically, $\geq 2\%$ difference between closely related species, as defined by high-quality morphological studies), and sequences from members of the same species should differ very little. One purpose of barcoding campaigns is to test whether the barcode gene displays this desired high interspecific difference and low intraspecific variation for a given taxon. Second, the flanking regions of barcode sequences should display very low sequence variation so that it is easy to amplify the barcode sequence using polymerase chain reaction (PCR). Part of the art of barcoding is to design 'universal' PCR primers that can be used on a wide range of taxa. A classic example is the Folmer primer pair (Folmer *et al.* 1994), which can amplify COI across large swathes of the Insecta.

The second technology used in metabarcoding is high-throughput sequencing. Standard Sanger technology is limited to sequencing a single gene from a single specimen in each run. High-throughput sequencers, in contrast, can separately sequence individual DNA molecules and thus accept mixtures of genes, specimens and species. Kircher & Kelso (2010) provide an entry to the technology, Glenn (2011) is an update, and there are many videos on YouTube. Readers should be aware that new machines and upgrades to those machines appear constantly and advance on several fronts, including total throughput (known as sequencing 'depth' or 'coverage'), sequence (or 'read') length and quality, cost and run times.

Metabarcoding thus uses universal PCR primers to mass-amplify a taxonomically informative gene from mass collections of organisms or from environmental DNA, and the prefix 'meta' thus refers to the collection of barcode genes. The PCR product (an 'amplicon') is sent to a high-throughput sequencer, and the output is a long list of DNA sequences. PCR and sequencing introduce errors into the sequences, which are removed or fixed on a computer. Next, because each individual of each species has contributed many DNA strands, each of which has then been copied many times by PCR, the output data set needs to be reduced by using a computer to cluster the sequences into 'operational taxonomic units,' or OTUs, each of which ideally should contain only the sequences from one species. Finally, a representative sequence is taken from each OTU and assigned a taxonomy using one or more of the databases listed above.

invoked the high plausibility of the taxonomies and the ecological patterns uncovered (e.g. Chariton *et al.* 2010; Fonseca *et al.* 2010; Nolte *et al.* 2010; Porazinska *et al.* 2010; Hajibabaei *et al.* 2011; Thomsen *et al.* 2011; Hiiesalu *et al.* 2012; Yoccoz *et al.* 2012; Baldwin *et al.* 2013).

In general, these studies have found that not every species is recovered from samples and that the ecological patterns do not perfectly match those found using standard data sets. Can these discrepancies be ignored? Are the metabarcode data sets in fact revealing higher resolution ecological patterns? Most importantly, *can the information that is recovered by metabarcoding be used to answer policy and management questions reliably*?

To answer these questions, we must compare the performance of metabarcode data sets against high-quality biodiversity data sets that have been collected to answer real policy questions. Only in this way can metabarcoding make the transition from a research technology to a tool for environmental management that can have legal weight, and for designing and validating coarser but more cost-effective biodiversity indicators.

We therefore compare metabarcoding (MBC) data sets against three large-scale, high-quality, species-level, standard (STD) biodiversity data sets collected for the purpose of answering policy questions in conservation biology. We ask whether MBC and STD data sets result in similar estimates of alpha- and beta-diversity patterns and, more importantly, in similar policy conclusions. Our MBC data sets were collected in parallel with STD biodiversity data sets that comprise a total of 55,813 designated indicator specimens expertly identified to species level using morphological characters (Table 1). The conservation applications tested here are (1) measuring the effects of climate change on species distributions, which is a proxy for both targeted and surveillance biodiversity monitoring, (2) ecological restoration and (3) systematic conservation planning.

MATERIALS AND METHODS

Biodiversity sampling

Our samples were collected in three biomes, subtropical forest (Ailaoshan, China), temperate woodland (Thetford, UK) and tropical rainforest (Danum Valley, Malaysia) (Table 1). For two locations, Ailaoshan and Thetford, we metabarcoded the entire samples ('supersets') from which the STD indicator taxa had been drawn. For Danum Valley, MBC samples were collected separately from, but in parallel with, the STD samples (Tables 2 and 3). Danum Valley's STD and MBC samples therefore are expected to exhibit low to no taxonomic overlap. Detailed descriptions of scientific motivations, study sites, and sampling and taxonomic protocols for the three locations are in Supporting Information section S1.

Designated STD taxa were chosen, as always, via a compromise between available taxonomic expertise and workload capacity and

Table 1 Location descriptions. The designated indicator taxa were identified to species or morphospecies in the standard (STD) biodiversity data sets

Location	Biome	Lat/Long	Application	Sampling methods	Designated indicator taxa	Number of specimens	Taxonomic effort (person-h)
Ailaoshan, Yunnan, China	Subtropical forest	N 24.283 E 101.257	Biological effects of climate change	Light traps	Moths	8,002	676
Thetford, Norfolk, United Kingdom	Temperate woodland	N 52.412 E 0.657	Restoration ecology	Pitfall traps	Ants, Spiders, Carabid beetles	17,498	496
Danum Valley, Sabah, Malaysia	Tropical rainforest	N 5.006 E 117.816	Systematic conservation planning	Malaise traps, mist nets, Winkler traps, baited pitfall traps	Birds, Ants, Dung beetles	30,313	24 (Birds) 819 (Ants) 490 (Dung beetles)

The number of specimens and the person-hours of expert taxonomic effort apply to the STD data sets only.

Table 2 Beta-diversity comparisons of Metabarcoding (MBC) and Standard (STD) data sets

Location	Location subset	Does MBC sample include STD?	MBC, non-singleton spp	STD, non-singleton spp	Mantel r	*P*	NMDS and Procrustes r	*P*
Ailaoshan	All sites, $n = 39$	Yes + residue	985 Lepidoptera 98% OTUs	546 moth species	0.714	0.001	0.767	0.001
Ailaoshan	All sites, $n = 39$	Yes + residue	628 Arthropoda 97% OTUs	546 moth species	0.630	0.001	0.839	0.001
Thetford	With Heath sites, n = 67	Yes + residue	284 Arthropoda 97% OTUs	106 ant, spider, and carabid beetle species	0.512	0.001	0.608	0.001
Thetford	Without Heath sites, n = 60	Yes + residue	284 Arthropoda 97% OTUs	106 ant, spider, and carabid beetle species	0.233	0.001	0.421	0.001
Danum Valley	With Oil Palm sites, n = 26	No	1245 Arthropoda 97% OTUs	181 ant species	0.577	0.001	0.784	0.001
Danum Valley	With Oil Palm sites, n = 28	No	1317 Arthropoda 97% OTUs	51 dung beetle species	0.584	0.001	0.779	0.001
Danum Valley	Without Oil Palm sites, n = 22	No	1124 Arthropoda 97% OTUs	170 ant species	0.006	0.490*	0.144	0.889
Danum Valley	Without Oil Palm sites, n = 24	No	1198 Arthropoda 97% OTUs	46 dung beetle species	0.083	0.112	0.272	0.365
Danum Valley	Without Oil Palm sites, n = 24	No	1198 Arthropoda 97% OTUs	66 bird species	0.072	0.160	0.332	0.164

*Mantel $r = 0.198$, $P = 0.012$ if high-prevalence species (present in more than 15 sites) are removed from both data sets.

For each location or location subset, Mantel and Procrustes tests are used to compare Jaccard community dissimilarities among the N census sites. Significant correlations indicate that MBC and STD data sets estimate beta diversity similarly. Procrustes tests used the ordinations in Figs 1 and 2. For the Ailaoshan and Thetford data sets, the same samples were used as input for the MBC and STD data sets, except that the STD data set includes only the indicator taxa while MBC sample uses the entire sample (indicators + 'residue').

Table 3 Alpha-diversity comparisons of MBC and STD data sets. (a) Species richness in the Ailaoshan MBC (Lepidoptera-only) and STD (moth) data sets are not significantly different using two of three incidence estimators: Chao2 (see table), Jackknife1 (MBC 1434.5 ± 85.9 SE vs. STD 1575.3 ± 59.4 SE, $P > 0.1$, Welch's *t*-test), bootstrap (MBC 1187.7 ± 42.6 vs. STD 1435.9 ± 39.0, $P < 0.001$). Only two butterflies were captured in the light traps. In Thetford, as expected, total arthropod OTU richness is significantly greater than the number of Ant + Spider + Carabid beetle species (shown in table). Note that a 98% Arthropoda OTU threshold would only increase this disparity. (b) Chao2 species richness estimates for the Danum Valley MBC (Arthropoda 97% OTUs) and STD (Ants, Birds, Dung beetles) data sets, at three logging levels. Ant and Dung beetle richness are highest in unlogged sites, while MBC and Bird richness are highest in the twice-logged sites

(a)	Metabarcoding, all spp	Chao2	Standard, all spp	Chao2	Welch's *t*-test *P*
Ailaoshan	1284 Lepidoptera 98% OTUs	1446.0 ± 24.7 SE	996 moth morphospecies	1546.3 ± 69.9 SE	$P > 0.1$
Thetford, with Heath sites, $n = 67$	286 Arthropoda 97% OTUs	286.0 ± 0.14 SE	125 ant, spider, and carabid beetle species	146.6 ± 9.6 SE	$P < 0.001$
Thetford, without Heath sites, $n = 60$	270 Arthropoda 97% OTUs	271.9 ± 1.6 SE	129 ant, spider, and carabid beetle species	145.4 ± 11.1 SE	$P < 0.001$

(b)	Taxon	Unlogged Chao2	Once-logged Chao2	Twice-logged Chao2
Danum Valley	MBC	1815	2094	<u>2239</u>
	Ants	<u>256</u>	192	211
	Dung beetles	<u>65</u>	45	48
	Birds	83	63	<u>92</u>

For clarity, richness estimates are rounded to the nearest species, and highest estimates are underlined.

Table 4 Sequence databases for DNA taxonomy, all accessed 17 March 2013.

Database	Website	Focal taxa	Genes
GenBank	www.ncbi.nlm.nih.gov/genbank	All	Repository for official BARCODE sequences, plus a catch-all database
DDBJ	www.ddbj.nig.ac.jp	All	
ENA	www.ebi.ac.uk/ena/	All	
BOLD	www.boldsystems.org	Animals, Plants, Fungi, Protists	COI, ITS, RbcL, MatK
Greengenes	greengenes.lbl.gov	Bacteria, Archaea	16S
Silva	www.arb-silva.de	Bacteria, Archaea, Eukarya	16S, 18S, 23S, 28S
Unite	unite.ut.ee	Fungi	ITS

which taxa are thought to be informative for the question at hand. The need to make this compromise can be considered a weakness of STD.

Ailaoshan

Using this data set, we ask whether it is possible to use MBC to monitor the effect of climate change on biological communities. An altitudinal transect with light-trap samples taken at 2000, 2200, 2400 and 2600 m above sea level, and at two strata (canopy, ground), provides a climate gradient. Moths were the designated STD indicators and were extracted (physically removed from the samples), sorted to morphospecies and identified to family. Most samples were split into an STD and an MBC portion, but when sample volumes were small, the moths were extracted, sorted for STD, and whole bodies or legs were placed back in the samples for MBC. The MBC data set thus comprised whole or half light-trap samples, depending on volume, and included all taxa. (Details in Supporting Information section S1.1).

Thetford

With this data set, we ask whether it is possible to identify the ecological restoration treatments that are most effective at converting grass-covered forest trackways into hospitable habitat for heathland-specialist arthropod species, the goal being to connect fragments within heathland areas (Pedley *et al.* 2013). Trackways were subjected to one of six disturbance treatments (ranging in severity from mowing to turf-stripping) and sampled with pitfall traps. Ants, spiders and carabid beetles were the designated STD indicators and were extracted and identified to species. In parallel, whole pitfall-trap samples, including legs of the STD taxa, were metabarcoded. (Details in Supporting Information section S1.2).

Danum Valley

Here, we ask whether MBC data sets contain useful information for systematic conservation planning. Edwards *et al.* (2011) have reported that selectively logged rainforest in Borneo maintains bird and dung-beetle species richness at levels comparable to unlogged forest (dung beetles are a mammal indicator). Importantly, the timber values of once-logged and twice-logged forests are 40% and 20%, respectively, of the values of unlogged forest, suggesting that a portion of land-acquisition budgets could be efficiently spent on conserving more and cheaper logged forest (Fisher *et al.* 2011). We surveyed unlogged, once-logged and twice-logged forest patches for three designated STD indicators, using mistnets for birds, pitfall traps for dung beetles and Winkler extractors for leaf-litter ants, all of which were identified to species or morphospecies. Ants and dung beetles were also sampled in oil palm plantations. In parallel, Malaise traps collected MBC samples on the same trails used for the STD samples, and whole samples were metabarcoded. (Details in Supporting Information section S1.3).

Sample preparation, PCR strategy and 454 pyrosequencing of COI amplicons

We prepared MBC samples by using two legs from all specimens equal to or larger than a honeybee and whole bodies of everything smaller, adding 4 mL Qiagen ATL buffer (Hilden, Germany) (20 mg/ml proteinase k = 9 : 1) per 1.0 g of sample, homogenising with sterile 0.25-inch ceramic spheres in a FastPrep-24® system (MP Biomedicals, Santa Ana, CA, USA) set on 5 m/s for 1 min at room temperature, incubating overnight at 56 °C, and using 10% of the lysed solution for genomic DNA extraction with the Qiagen DNeasy Blood & Tissue Kit, using no more than 900 μL per spin column. The quantity and quality of purified DNA was assessed using the Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). Samples were PCR amplified using the degenerate primers, *Fol-degen-for* 5′-TCNACNAAYCAYAARRAYA-TYGG-3′ and *Fol-degen-rev* 5′-TANACYTCNGGRTGNCC-RAAR-AAYCA-3′. The standard Roche A-adaptor and a unique 10 bp MID (Multiplex IDentifier) tag for each sample (within collection) were attached to the forward primer. Each sample was amplified in three independent reactions and pooled. PCRs were performed in 20 μL reaction volumes containing 2 μL of 10 × buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.4 μM each primer, 0.6 U HotStart Taq DNA polymerase (TaKaRa Biosystems, Ohtsu, Japan), and approximately 60 ng of pooled genomic DNA. We used a touchdown thermocycling profile of 95 °C for 2 min; 11 cycles of 95 °C for 15 s; 51 °C for 30 s; 72 °C for 3 min, decreasing the annealing temperature by 1 degree every cycle; then 17 cycles of 95 °C for 15 s, 41 °C for 30 s, 72 °C for 3 min and a final extension of 72 °C for 10 min. We used non-proofreading Taq and fewer, longer cycles to reduce chimera production (Lenz & Becker 2008; Yu *et al.* 2012). For pyrosequencing, PCR products were gel-purified by using a Qiagen QIAquick PCR purification kit, quantified using the Quant-iT PicoGreen dsDNA Assay kit (Invitrogen, Grand Island, New York, USA), pooled and A-amplicon-sequenced on a Roche GS FLX at the Kunming Institute of Zoology. Further details are provided in Yu *et al.* (2012). The 39 Ailaoshan samples were sequenced on four 1/8 regions, producing 370 923 raw reads and 262 432 post-quality-control (QC) reads (mean read length 248 bp). The 68 Thetford samples were sequenced on four 1/16 regions, producing 71 661 raw reads and 45 621 post-QC reads (413 bp). The 56 Danum Valley samples were individually extracted and amplified, and then pooled within transect (2 per transect) for pyrosequencing on two 1/4 regions, producing 375 925 raw reads and 297 171 post-QC (445 bp). We did not rarefy these data sets to equalise read numbers across samples because (1) there is a high ratio of read number to species richness, relative to bacterial samples, meaning that we likely have covered most or all extractable arthropod biodiversity with our samples, (2) we know that some taxa are less likely to amplify at high read numbers than are other taxa, such as Hymenoptera (Yu *et al.* 2012), and rarefaction is inherently more likely to remove species represented by few reads and thus might introduce taxonomic bias.

Bioinformatic analysis

We followed an experimentally validated pipeline (Yu *et al.* 2012) to denoise and cluster the reads into Operational Taxonomic Units (OTUs). *Quality control*: Header sequences and low-quality reads were removed from the raw output in the QIIME 1.5.0 environment (*split_libraries.py*: -l 100 -L 700 -H 9 -M 2 -b 10) (Caporaso *et al.* 2010b). *Denoising and chimera removal*: PyNAST (Caporaso *et al.* 2010a) was used to align reads against a high-quality, aligned data set of Arthropoda sequences (Yu *et al.* 2012) at a minimum similarity of 60%, and sequences that failed to align were removed. The remaining sequences were clustered at 99% similarity with USEARCH (Edgar 2010), a consensus sequence was chosen for each cluster, and the UCHIME function was used to perform *de novo* chimera detection and removal. A clustering step is required for chimera detection because chimeric reads are expected to be rare and thus belong to small clusters only. The final denoising step used MACSE (Ranwez *et al.* 2011), which aligns at the amino acid level to high-quality reference sequences and uses any stop codons in COI to infer frameshift mutations caused by homopolymers. We removed any sequences < 100 bp. *OTU-picking and Taxonomic assignment*: To reduce total computation time, sequences were first chain-clustered at 99% similarity using DNACLUST (Ghodsi *et al.* 2011) and then at 97% using CROP (Hao *et al.* 2011). OTUs were assigned taxonomies using SAP (Munch *et al.* 2008), keeping only taxonomic levels for which the posterior probability was > 80%. OTUs containing only one read or assigned to non-arthropod taxa were removed. In the Ailaoshan data set, Lepidoptera-assigned OTUs were extracted, expanded and re-clustered at 98% similarity to increase our power to differentiate closely related species. Computations were performed on a combination of Apple iMacs and a Linux computing cluster at the University of East Anglia (rscs.uea.ac.uk/high-performance-computing, accessed 18 May 2013). Sequence data are deposited at datadryad.org (doi: 10.5061/dryad.t3v71) and in GENBANK's Short Read Archive (Accession numbers in Supporting Information S6).

Statistical analysis

Most analyses were performed using *R* (R Core Team 2012), *vegan* (Oksanen *et al.* 2012), and *mvabund* (Warton *et al.* 2012). An example *R* script and input data sets are deposited at datadryad.org (doi: 10.5061/dryad.t3v71). For each of the three locations, we have an STD and an MBC Species/OTU X Sample table, plus associated environmental variables. We removed singleton OTUs and Species and converted MBC read numbers to presence/absence (Yu *et al.* 2012).

To visualise the effects of environmental treatment levels on community compositions, we used non-metric multidimensional scaling (NMDS) ordination of Jaccard dissimilarity matrices (Fig. 1), which were created with *vegan's vegdist, metaMDS*, *plot* and *ordiellipse* functions. To test whether the effect sizes of the environmental treatment levels were similar across the STD and MBC data sets, we used *vegan's mantel* and *protest* correlation tests (Table 2). To compare species richness, we used incidence coverage estimators, which were calculated with *vegan's specpool* function (Table 3).

For hypothesis testing, we used *mvabund* to test the effects of environmental predictors on community composition. *mvabund* is a multivariate implementation of generalised linear models, and, unlike dissimilarity-matrix-based methods, *mvabund* does not confound location with dispersion effects, which can inflate type 1 and 2 errors (Warton *et al.* 2012). The *summary.manyglm* function in *mvabund* was used for treatment contrasts. That is, we tested for significant

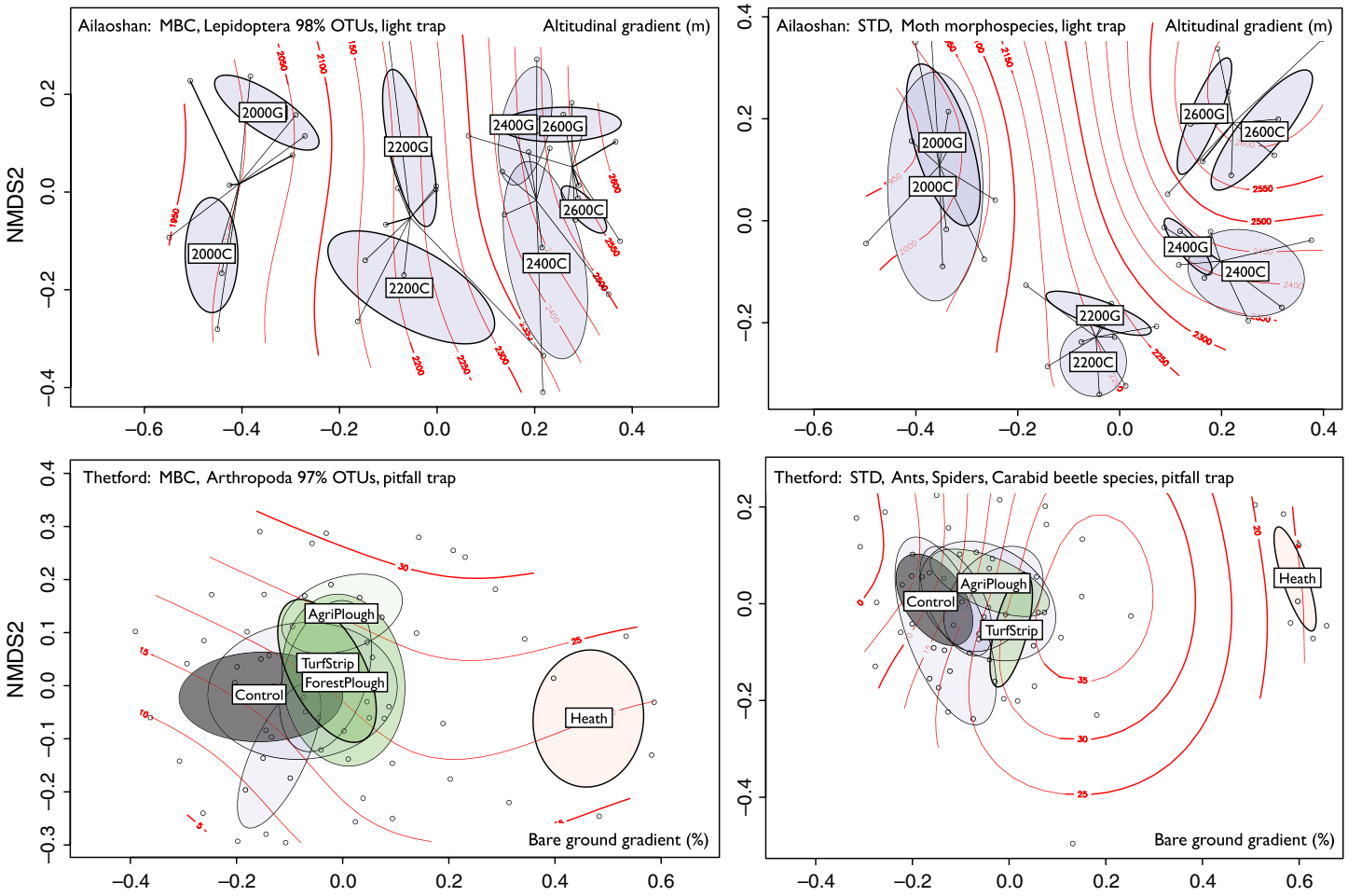

Figure 1 Non-metric multidimensional scaling (NMDS) ordinations. Points are census sites, and coloured ellipses are 95% confidence intervals of species centroids for each treatment level ['ordiellipses' (Oksanen *et al.* 2012)]. These ordinations are for visualisation; all statistical tests of treatment effects are conducted using *mvabund* (see text for details). (a, b**)**. Ailaoshan. Altitude and Strata (canopy, ground) effects on moth communities. Sites within the same altitude are connected by line segments, with ellipses drawn for each combination of altitude and stratum. An NMDS ordination of the metabarcoding (MBC) Arthropoda 97% operational taxonomic units (OTUs) data set is in Supporting Information section S2. (c, d). Thetford. Restoration treatment effects on MBC (Arthropoda 97% OTUs) and STD (ants, spiders, carabid beetles) communities. Green ellipses indicate treatments that significantly shifted community composition away from the undisturbed Control (black) sites towards the target Heathland habitat (red). Blue ellipses indicate treatments that are not significantly different from Control sites (left unlabelled for clarity).

differences of disturbance (Thetford) and logging (Danum Valley) treatment levels on community composition, relative to controls, and we corrected for multiple tests using the *p.adjust(method='fdr')* function in R's base package.

For the conservation planning application, we used RSW2 (Arponen *et al.* 2005) with default parameter settings, 10 000 runs, two replicates, and equally valued species, to choose the set of sites that maximised species coverage under each budget. The STD data set included only Birds and Dung beetles because the Ants data set was incomplete, due to heavy rains that prevented collection at two transects. To test the degree to which RSW2 outputs are correlated between the STD and MBC data sets over and above similarities created by pure budget effects, we devised a Monte Carlo test that randomly selected site subsets 10 000 times, constrained by each of the six budgets. The null probability of matching RSW2 outputs is given by the proportion of runs that have as many or more matches as the RSW2 solution. An *R* script and an example data set are deposited at datadryad.org (doi: 10.5061/dryad.t3v71).

RESULTS

Ailaoshan

We generated two MBC data sets. The first included only Lepidoptera OTUs, clustered at 98%, and the second included all Arthropoda-OTUs clustered at a 97% similarity threshold. The Lepidoptera data set allows direct comparison with the STD moth data set (only two butterflies were collected), while the arthropod data set takes advantage of MBC's taxonomic comprehensiveness.

NMDS ordinations reveal clear and very similar community compositional differences across Altitude and Stratum levels in the STD and MBC data sets (i.e. the 'effect sizes' of the environmental variables are substantial and similar across data sets). As expected, beta diversity structures in the Lepidoptera-only and in the all-Arthropoda MBC data sets are both highly significantly correlated with the STD moth data set, as tested by Mantel tests and by Procrustes

analysis on the NMDS ordinations (Table 2; Fig. 1a,b; Supporting Information section S2).

Also, total lepidopteran species richness, as estimated by two of three incidence-based estimators, was not significantly different across the MBC and STD data sets (Table 3a).

Consistent with the ordinations, the *anova.manyglm* test in *mvabund* found that the Altitude and Stratum predictors both had highly significant main effects on community composition in both the MBC and STD data sets (all $P = 0.001$). Interaction effects were non-significant (MBC: $P = 0.482$; STD: $P = 0.542$) (Fig. 1a,b) (see Supporting Information section S3 for *mvabund* statistical details).

In summary, the MBC and STD data sets detect the same changes in community composition across an altitudinal and a micro-habitat gradient. Both data sets also return similar estimates of total species richness.

Thetford

We generated an MBC data set consisting of Arthropoda-OTUs clustered at 97% similarity.

NMDS ordinations reveal that the eight treatment levels (control, six disturbance levels, heathland) resulted in similar community compositional differences (i.e. effect sizes) across the eight treatment levels in the STD and MBC data sets (Fig. 1c,d). These community responses are significantly correlated across the MBC and STD data sets, as shown by Mantel and Procrustes tests (Table 2). The statistical significance of the whole-data set correlations is driven in part by the influential heathland sites, but a second round of tests excluding the heathland sites remains statistically highly significant (Table 2).

Because we metabarcoded all taxa in the pitfall traps, the estimated species richness of the MBC Arthropoda-OTUs is unsurprisingly higher than that of the ant + spider + carabid STD data set (Table 3a). However, the large number of sampled sites in the Thetford data set allows us to use incidence-coverage estimators, and we find that MBC and STD can both detect when a restoration treatment results in high or low species richness. Four of five species richness estimators are significantly positively correlated across treatment levels (Species observed: $\rho_{\text{Spearman's coefficient}} = 0.886$, $P = 0.003$; Chao2: $\rho = 0.571$, $P = 0.151$; Jackknife1: $\rho = 0.833$, $P = 0.015$; Jackknife2: $\rho = 0.786$, $P = 0.028$; Bootstrap: $\rho = 0.905$, $P = 0.005$; *specpool* function in *vegan* (Oksanen *et al.* 2012) (see Supporting Information section S4 for scatterplots).

The MBC and STD data sets can also identify which treatments are most effective for restoring trackway into habitats that can support heathland arthropods (Fig. 1 c,d). In the MBC data set, the three heaviest disturbance treatments, AgriPlough ($P = 0.018$), TurfStrip ($P = 0.016$), and ForestPlough ($P = 0.011$) all resulted in communities that significantly diverge from the Control sites in the direction of the target Heath habitat (treatment contrasts conducted with *summary.manyglm* in *mvabund*). In the STD data set, the two heaviest disturbance treatments, AgriPlough ($P = 0.038$) and TurfStrip ($P = 0.038$), were identified as being different from the control sites, but the more moderate ForestPlough treatment diverged weakly from the control ($P = 0.153$) (see Supporting Information section S3 for *mvabund* statistical details), which might be due to lower statistical power in the smaller STD data set or a lack of response by the ant species, which dominate the data set.

In summary, the MBC and STD data sets return correlated estimates of species richness. Both data sets also identify the heavier disturbance treatments as being more effective at converting trackways into hospitable corridors for heathland arthropods.

Danum Valley

We generated an MBC data set consisting of Arthropoda-OTUs clustered at 97% similarity.

Unlike the two previous examples, the MBC and STD samples are taxonomically distinct. Malaise traps (MBC) capture mainly flying insects, whereas birds and dung beetles (STD) are indicators of vertebrate communities. Nonetheless, all the data sets with oil-palm samples (MBC, ant, and dung beetles) successfully reveal that oil-palm and forest sites have very different species compositions (Table 2; birds were not mist-netted in oil palm). On the other hand, species richness estimates in MBC and STD data sets are uncorrelated across the four habitats. Twice-logged forests host more bird species and Arthropoda-OTUs, while unlogged forests host more ant and dung-beetle species (Table 3b).

The more policy-relevant challenge is to differentiate among just the three logging levels in the forest sites (Fig. 2a–d). We observe moderate agreement within the STD data sets and between the MBC and STD data sets. Birds and dung beetles both differentiate unlogged forests from twice-logged (birds: $P = 0.024$, dung beetles: $P = 0.007$, *summary.manyglm*) and once-logged ($P = 0.024$, $P = 0.026$) forests. Ants fail altogether to differentiate the three logging levels ($P > 0.10$). The MBC data set lies in the middle, differentiating unlogged forests from twice-logged ($P = 0.014$), but not from once-logged ($P = 0.39$) forests (see Supporting Information section S3 for *mvabund* statistical details).

In short, communities of birds and dung beetles (the latter are mammal indicators) seem to respond more sensitively to logging, relative to arthropods (MBC) and ants (STD). Thus, the differences amongst data sets might indicate that vertebrates are more sensitive to logging.

As a statistical aside, note that the *mvabund* test for the dung beetle data set seems to contradict the NMDS visualisation, in that once-logged and unlogged forests have overlapping species centroids (Fig. 2d), whereas *mvabund* detects that dung beetles differentiate once-logged from unlogged forests. This disagreement could result from suboptimal solution finding in the NMDS ordination and/or from the community-level heteroscedasticity that causes Type 1 and 2 errors in dissimilarity-matrix-based analyses (Warton *et al.* 2012).

Given these differences and similarities among the four data sets, is it still possible to come to similar *policy conclusions* regarding the conservation value of selectively logged rainforest? Because the forests differ in both richness (Table 3b) and composition (Fig. 2a–d), we treat this question as a problem of systematic conservation planning, and we use the software package RSW2 (Arponen *et al.* 2005) to maximise total species coverage for subsets of the 24 census sites acquired under various budgets. Both the MBC data set and a combined dung beetle+bird STD data set (thus, weighted towards vertebrates) return very similar acquisition strategies that are weighted towards the cheap but still species-rich twice-logged forest. As budgets increase, once- and unlogged sites are acquired to complement the twice-logged sites (Fig. 3).

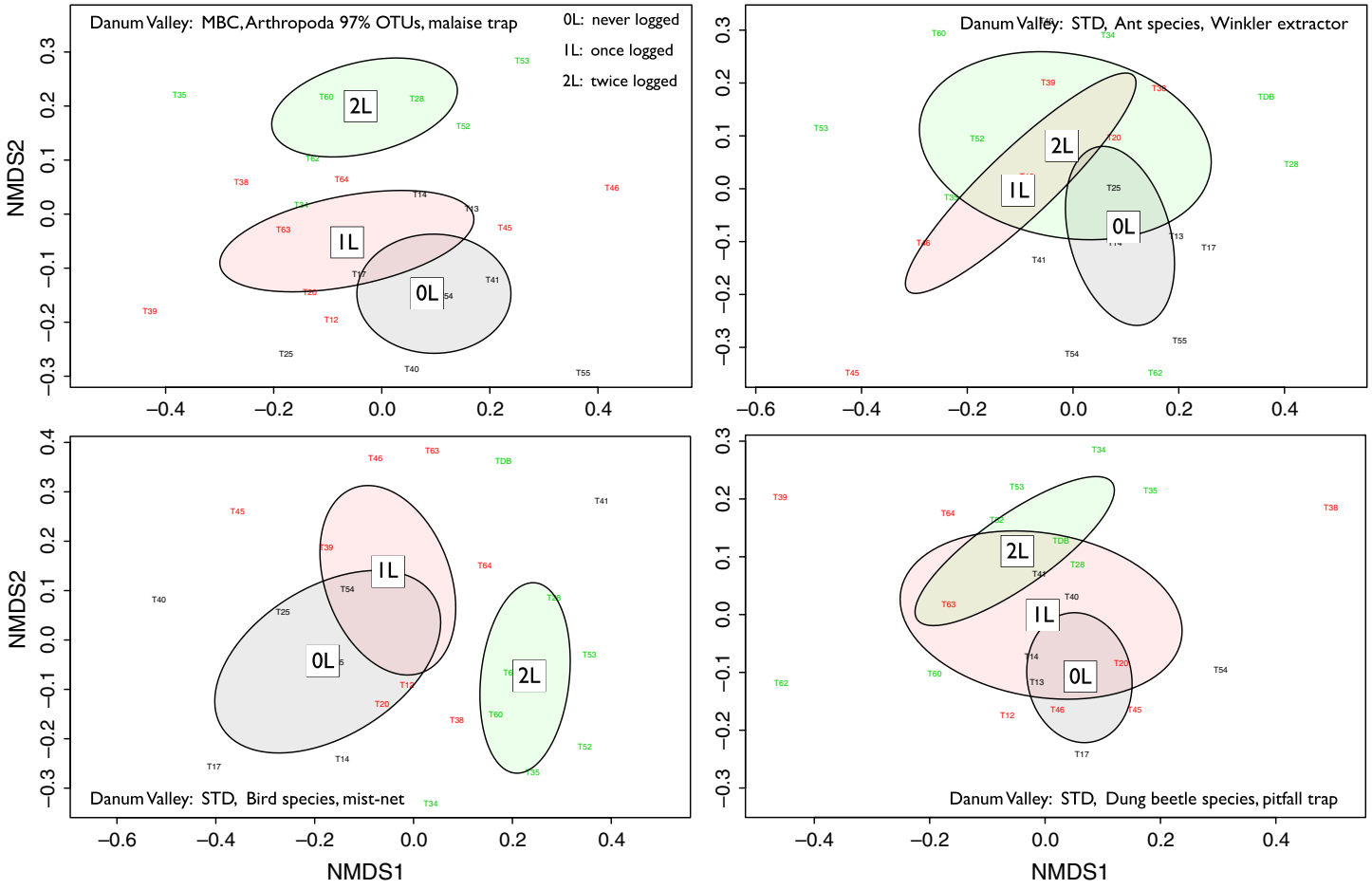

Figure 2 Non-metric multidimensional scaling (NMDS) ordinations. **A-D.** Danum Valley. Logging effects on metabarcoding (Arthropoda 97% operational taxonomic units) and STD (ants, birds and dung beetles) communities. The very dissimilar Oil-palm sites are omitted to allow logging effects to be clearly differentiated.

Naturally, because we imposed budget constraints, lower cost sites are more likely to be chosen over higher cost sites, which by itself generates a trivial similarity between STD and MBC RSW2 outputs. To test whether the community compositions of the 24 sample sites also contribute importantly to the acquisition choices, we carried out a Monte Carlo test, which found that the vertebrate-biased STD and the arthropod-only MBC data sets result in acquisition strategies that are significantly or marginally significantly more similar than expected from budget effects alone (Fig. 3) (see histograms of the Monte Carlo test output in Supporting Information section S5).

DISCUSSION

In summary, we show that metabarcoding is a reliable method for recovering alpha- and beta-diversity information from large-scale, field-collected data sets (Tables 2 and 3; Figs 1 and 2; Supporting Information section S2, S3, S4), even when tested against the highest quality STD data sets that can reasonably expected to be gathered under normal financial and time constraints. Reassuringly, Mantel and Procrustes correlation coefficients are highest when the STD and MBC data sets are focused on the same taxon subset (Ailaoshan moths, $r = 0.714$ & 0.767), mostly high to medium when the MBC data set is a superset of the STD data set (Ailaoshan Arthropoda: $r = 0.630$, 0.839; Thetford Arthropoda: $r = 0.233$–0.608), and low and non-significant only in Danum Valley, where STD and MBC samples did not overlap taxonomically (MBC used Malaise traps, STD used pitfall-trapped ants and dung beetles, plus birds) (Table 2).

The STD and MBC data sets also return very similar statistical models and policy conclusions. This is the first demonstration that MBC is a reliable source of biodiversity information for policymaking.

In Ailaoshan, both the STD & MBC data sets allow the detection of highly significant main effects of Altitude and Stratum and fail to find interaction effects (Fig. 1a,b; Supporting Information section S2, S3). Given this demonstrated sensitivity of MBC to changes in arthropod community composition with altitude, we propose that MBC can be used to monitor how communities shift in response to environmental change, such as how higher altitude (and -latitude) communities are expected to become more similar to lower altitude (and -latitude) communities with global warming.

In Thetford, the MBC and STD data sets both reveal that the highest-disturbance treatments show the biggest shifts away from the control sites and towards the target heathland habitat (Fig. 1c,d; Supporting Information section S3). We therefore propose that MBC can be used to monitor responses to restoration experiments.

Finally, in Danum Valley, MBC and STD data sets return similar acquisition strategies for systematic conservation planning (Fig. 3).

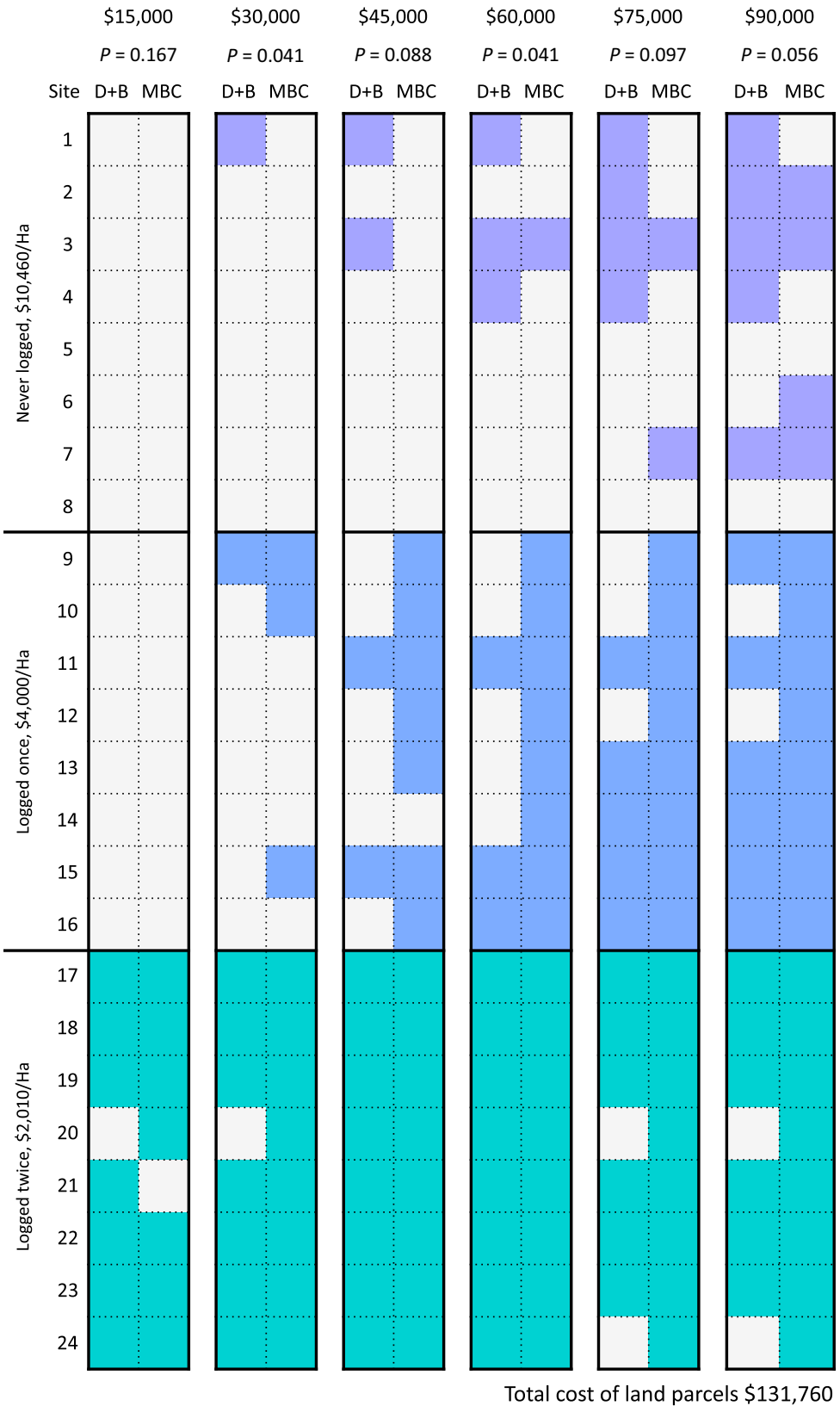

Figure 3 Optimal site selection in Danum Valley using the conservation planning software package RSW2. STD and metabarcoding data sets are compared at six acquisition budgets. The STD data set pools dung beetles+birds (D+B). Acquired sites at a given budget are indicated by a coloured square (twice-logged: green; once-logged: blue; never logged: purple). *P*-values under each budget amount indicate the probability that the proportion of sites acquired by both data sets can be accounted for by budget-constrained-chance alone (see also Supporting Information section S5 for histograms of the Monte Carlo outputs). A *P*-value <0.05 thus indicates that community compositions significantly contribute to site selections.

MBC can therefore be an efficient way to gather new conservation planning data, and to supplement any existing STD data, such as bird and mammal distributions.

Cost-effectiveness

For the 134 samples in the three STD data sets, a total of 2,505 person-hours of taxonomic expertise were expended for specimen identification (Table 1). In contrast, the active workload for the MBC data set (from samples to OTUs + taxonomies) was four times smaller, at 645 person-hours (571 person-hours for DNA extraction of 163 samples, which includes two samples per transect in Danum Valley, 54 for PCR and gel purification, and 20 for bioinformatic analysis). A further 520 hours were expended in the background (180 h for pyrosequencing and 340 h of computer time). Even this contrast underestimates the efficiency of metabarcoding because (1) the MBC data sets include all taxa, and (2) laboratory skills are much more abundant than is taxonomic expertise, meaning fewer delays before sample processing. Thus, a standard molecular laboratory with just a few staff can process many hundreds of whole samples annually, from anywhere in the world, a rate and breadth of data production that is inconceivable using the standard approach. We estimate a monetary cost of US$240–415 per sample, with the variation driven by labour and sequencing costs, the latter of which is declining rapidly. Hajibabaei *et al.* (2012) have recently proposed that it is possible to extract representative DNA from the ethanol used to preserve samples. If this can be validated for large-scale work, consumables and labour costs would decrease as well. Note that metabarcoding costs increment by sample, while standard biodiversity costs increment by specimen, which is why standard biodiversity censuses limit themselves to indicator taxa.

We deliberately do not present our costs for the STD data sets because they would be misleading. Biologists producing STD data sets for research are an inelastic and heterogeneous resource and derive personal utility on top of salaries. Someone wanting to contract for STD data sets could not budget on the basis of our prorated salary costs. Instead, they would need to hire whatever expertise is available, and this often means expensive, short-term, narrow-scope studies of unknown quality that cannot be standardised across landscapes or over time. We have provided time budgets in Table 1, which can be used to size STD contracts.

Advantages of standard biodiversity data sets

STD data sets currently have two important advantages over MBC data sets. First, STD data sets provide within-sample abundance information, which can be used for estimating local species diversities and inferring population dynamics. In contrast, while the number of sequences per OTU could be taken as an estimator of per species biomass, in practice, there is an unknown and probably non-trivial amount of error introduced by the vagaries of PCR and other laboratory and bioinformatic steps, including our cost-saving step of using only the legs of large specimens. Yu *et al.* (2012) found that 24% of the species in their constructed samples were not detected ('dropout') and that read numbers did not correlate with abundance in a preliminary experiment involving experimentally varied moth numbers. Thus, Yu *et al.* (2012) recommended that MBC data sets should conservatively be converted to presence/absence; abundance and species richness can be estimated using incidence-coverage estimators (Table 3). On the other hand, note that DNA barcoding regularly uncovers morphologically cryptic species complexes (Janzen *et al.* 2005), which is the parallel of dropout in *STD* data sets.

A second potential advantage of STD data sets is greater taxonomic resolution, at least for some locales. Using the conservative SAP assignment method (Munch *et al.* 2008), we can assign almost all OTUs to order level but only ~ 15% of OTUs to family, genus, and species level (Yu *et al.* 2012), whereas, if the taxonomic expertise is available and the fauna is known, higher resolution is possible for STD collections. For instance, all species in the Thetford, UK STD data set were assigned Latin binomials. Advantages of greater taxonomic resolution are the assignment of ecological function and the detection of species of economic or cultural importance. However, as sequencing technology and bioinformatic software advance, we expect to be able to recover longer and more accurate sequences, which will allow higher confidence in and greater resolution of taxonomic assignments. Equally important will be the continued growth of the Barcode of Life Database and others (Box 1). It is only through the continued generation and maintenance of individually barcoded and curated specimens in museum collections that we will be able to link metabarcoding sequences to our vast storehouse of functional biological knowledge (Janzen *et al.* 2005). If metabarcoding is adopted by commercial or state users for biodiversity monitoring, such users could justify and provide continued funding for alpha taxonomy and the generation of high-quality barcode databases.

Benefits of metabarcoding

There are potentially several benefits that metabarcoding could bring to biodiversity conservation and environmental management. First, metabarcoding frees research and management to move away from *biodiversity indicators* and towards direct measurement of *total biodiversity* (Lindenmayer & Likens 2011). Indicators have been criticised as an inherently problematic approach to biodiversity measurement because their taxonomic representativeness (see Introduction) and robustness as measures of policy success are questionable (Lindenmayer & Likens 2011; Dolman *et al.* 2012; Nicholson *et al.* 2012) and because once an indicator is used as a policy target, the potential for manipulation can bias incentives and thereby cause the indicator to lose value as an indicator (Newton 2011). In contrast, metabarcoding generates standardised and broad measures of biodiversity, as we show here, and the possibility exists to calibrate remote-sensing data and to test the validity of and to refine existing biodiversity indicators, which is more difficult and costly with STD data sets. Recall also that metabarcoding can be used to census plant and vertebrate species via environmental DNA and, potentially, via parasites carrying host tissue (Rougerie *et al.* 2010; Schnell *et al.* 2012; Calvignac-Spencer *et al.* 2013) (Box 2).

The gains in cost-effectiveness and comprehensiveness made possible by metabarcoding make it easier to justify large-scale surveillance of biodiversity trends (Wintle *et al.* 2010; Possingham *et al.* 2012). Furthermore, the ability to monitor rapidly, reliably, comprehensively, cheaply, and in a third-party-verifiable way may increase the effectiveness of institutions that have been designed to conserve biodiversity (Zabel & Roe 2009; Baird & Hajibabaei 2012). The European Union, e.g. spends five billion euros annually on agri-environment programmes but struggles to determine which interventions result in cost-effective, sustainable and general conserva-

Box 2 A research agenda for metabarcoding.

The field of metabarcoding is advancing rapidly, and even the name of the method has not yet settled. Other names include (eco)metagenetics (Porazinska *et al.* 2010), environmental barcoding (Hajibabaei *et al.* 2011), biomonitoring 2.0 (Baird & Hajibabaei 2012), ecogenomics (ecogenomic.org/whatisecogenomics, accessed 29 June 2013), environmental sequencing (Fonseca *et al.* 2010) and simply bulk sequencing. Strictly speaking, the suffix 'omics' should be used only when all the DNA in a sample is sequenced, whereas barcoding and 'genetics' should be used when specific genes are sequenced ('targeted or amplicon sequencing').

Naturally, metabarcode data sets are subject to error and loss of information, so most research effort to date has been to validate metabarcoding against standard biodiversity censuses (see Introduction), and to develop more efficient and reliable pipelines that take advantage of advances in sequencing technology (e.g. Zhou *et al.* 2013). Another focus has been devising clever ways to collect the DNA of difficult-to-trap taxa: water, soil, pollen traps, faeces and parasites (Goldberg *et al.* 2011; Jerde *et al.* 2011; Pompanon *et al.* 2011; Thomsen *et al.* 2011; Andersen *et al.* 2012; Folloni *et al.* 2012; Hiiesalu *et al.* 2012; Schnell *et al.* 2012; Thomsen *et al.* 2012; Yoccoz *et al.* 2012; Calvignac-Spencer *et al.* 2013; Takahara *et al.* 2013). We expect both of these areas to continue to consume research effort.

In addition, we see the following three directions as especially important if metabarcoding is to bridge the science-practitioner divide:

(1) Developing statistical and laboratory methods to allow robust inference of species abundances in samples and across landscapes. Related to this is the development of PCR-free methods that reduce read-number biases and allow the detection of taxa that do not amplify well, such as the Hymenoptera (Yu *et al.* 2012).

(2) Robust methods of taxonomic assignment and phylogenetic placement, with confidence estimates at each taxonomic level, while minimising false-positive assignments (Matsen *et al.* 2010; Zhang *et al.* 2012).

(3) Deeper connection with the end-users of biodiversity data (Cook *et al.* 2013), including the development of chain-of-evidence and bioinformatic-reporting protocols to increase the credibility of the data.

tion gains (Kleijn *et al.* 2011). Metabarcoding can provide the high-volume data needed to measure local- and landscape-scale responses to agri-environment interventions, although work remains to translate such data to measures of abundance, and then to population viability (Box 2).

More generally, biodiversity-offset, environmental certification, and payments for environmental services schemes are beset with 'asymmetric-information' problems, which, at best, waste money, and, at worst, lead to biodiversity loss and deter attempts to implement conservation actions in the first place (Ferraro & Pattanayak 2006; Zabel & Roe 2009; Bekessy *et al.* 2010; Ferraro 2011; Kinzig *et al.* 2011; Newton 2011; Bottrill & Pressey 2012; Meijaard & Sheil 2012). The effectiveness of such contracts for biodiversity conservation and management might be increased by characterising the biodiversity endowments of potential land sellers and by allowing auditors, managers and consumers to condition payments, in part, on biodiversity outcomes, as well as on prescribed actions (Ferraro *et al.* 2005; Ferraro 2008; Wunder 2008; Zabel & Roe 2009; Yu 2010; Ferraro 2011; Gibbons *et al.* 2011; Meijaard & Sheil 2012). If chain-of-evidence and reporting protocols can be established (Box 2), metabarcoding provides one way to uncover the relevant information.

ACKNOWLEDGEMENTS

We thank Yang Yahan, Alice Wang, Vincent Moulton, David Warton and Wadud Miah for support and advice and to Ding Zhaoli for sequencing. LA, YT, AN and RK were supported by the Queensland-Chinese Academy of Sciences (QCAS) Biotechnology Fund (GJHZ1130) and Griffith University. DPE was supported by a STEP fellowship at Princeton University. SP was supported by the Natural Environment Research Council, Forestry Commission, Norfolk Biodiversity Information Service and Suffolk Biodiversity Partnership. Additional support for DPE, PW, FAE, THL and WHH was provided by a grant from the High Meadows Foundation to DSW. YQJ, XYW and DWY were supported by Yunnan Province (20080A001), the Chinese Academy of Sciences (0902281081, KSCX2-YW-Z-1027), the National Natural Science Foundation of China (31170498), the Ministry of Science and Technology of China (2012FY110800), the University of East Anglia, and the State Key Laboratory of Genetic Resources and Evolution at the Kunming Institute of Zoology.

AUTHOR CONTRIBUTIONS

YJ and DWY designed the study. YJ produced the metabarcoding data sets. LA, SP, DPE, YT, AN, RK, PD, PW, FAE, THL, WWH, SB, KCH, DSW and CB produced the standard biodiversity data sets, led by LA, SP, and DPE. YJ, ML and DWY conducted the bioinformatic analyses, for which XW wrote computer code. DWY conducted the statistical analyses. TL conducted the RSW2 resampling test. DWY wrote the first draft. YJ, LA, SP, DPE, YT, AN, RK, PD, PW, FAE, THL, DSW, TL, ML and BCE discussed the article and made revisions.

REFERENCES

Andelman, S.J. & Fagan, W.F. (2000). Umbrellas and flagships: efficient conservation surrogates or expensive mistakes? *Proc. Natl. Acad. Sci. USA*, 97, 5954–5959.

Andersen, K., Bird, K.L., Rasmussen, M., Haile, J., Breuning-Madsen, H., Kjaer, K.H. *et al.* (2012). Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Mol. Ecol.*, 21, 1966–1979.

Arponen, A., Heikkinen, R.K., Thomas, C.D. & Moilanen, A. (2005). The value of biodiversity in reserve selection: representation, species weighting, and benefit functions. *Conserv. Biol.*, 19, 2009–2014.

Asner, G.P., Powell, G.V.N., Mascaro, J., Knapp, D.E., Clark, J.K., Jacobson, J. *et al.* (2010). High-resolution forest carbon stocks and emissions in the Amazon. *Proc. Natl. Acad. Sci. USA*, 107, 16738–16742.

Baird, D.J. & Hajibabaei, M. (2012). Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Mol. Ecol.*, 21, 2039–2044.

Baldwin, D.S., Colloff, M.J., Rees, G.N., Chariton, A.A., Watson, G.O., Court, L.N. *et al.* (2013). Impacts of inundation and drought on eukaryote biodiversity in semi-arid floodplain soils. *Mol. Ecol.*, 22, 1746–1758.

Bekessy, S.A., Wintle, B.A., Lindenmayer, D.B., Mccarthy, M.A., Colyvan, M., Burgman, M.A. *et al.* (2010). The biodiversity bank cannot be a lending bank. *Conserv. Lett.*, 3, 151–158.

Bik, H.M., Porazinska, D.L., Creer, S., Caporaso, J.G., Knight, R. & Thomas, W.K. (2012). Sequencing our way towards understanding global eukaryotic biodiversity. *Trends Ecol. Evol.*, 27, 233–243.

Bottrill, M.C. & Pressey, R.L. (2012). The effectiveness and evaluation of conservation planning. *Conserv. Lett.*, 5, 407–420.

Calvignac-Spencer, S., Merkel, K., Kutzner, N., Kühl, H., Boesch, C., Kappeler, P.M. *et al.* (2013). Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Mol. Ecol.*, 22, 915–924.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L. & Knight, R. (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*, 26, 266–267.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K. *et al.* (2010b). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*, 7, 335–336.

Chariton, A.A., Court, L.N., Hartley, D.M., Colloff, M.J. & Hardy, C.M. (2010). Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. *Front. Ecol. Environ.*, 8, 233–238.

Cook, C.N., Mascia, M.B., Schwartz, M.W., Possingham, H.P. & Fuller, R.A. (2013). Achieving conservation science that bridges the knowledge-action boundary. *Conserv. Biol.*, 27, 669–678.

Cushman, S.A., McKelvey, K.S., Noon, B.R. & McGarigal, K. (2010). Use of abundance of one species as a surrogate for abundance of others. *Conserv. Biol.*, 24, 830–840.

Dolman, P.M., Panter, C.J. & Mossman, H.L. (2012). The biodiversity audit approach challenges regional priorities and identifies a mismatch in conservation. *J. Appl. Ecol.*, 49, 986–997.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26, 2460–2461.

Edwards, D.P., Larsen, T.H., Docherty, T.D.S., Ansell, F.A., Hsu, W.W., Derhé, M.A. *et al.* (2011). Degraded lands worth protecting: the biological importance of Southeast Asia's repeatedly logged forests. *Proc. R. Soc. Lond. B*, 278, 82–90.

Ferraro, P. (2008). Asymmetric information and contract design for payments for environmental services. *Ecol. Econ.*, 65, 810–821.

Ferraro, P.J. (2011). The future of payments for environmental services. *Conserv. Biol.*, 25, 1134–1138.

Ferraro, P.J. & Pattanayak, S.K. (2006). Money for nothing? a call for empirical evaluation of biodiversity conservation investments. *PLoS Biol.*, 4, 482–488.

Ferraro, P.J., Uchida, T. & Conrad, J.M. (2005). Price premiums for eco-friendly commodities: are 'green' markets the best way to protect endangered ecosystems? *Environ. Resour. Econ.*, 32, 419–438.

Fisher, B., Edwards, D.P., Larsen, T.H., Ansell, F.A., Hsu, W.W., Roberts, C.S. *et al.* (2011). Cost-effective conservation: calculating biodiversity and logging trade-offs in Southeast Asia. *Conserv. Lett.*, 4, 443–450.

Folloni, S., Kagkli, D.-M., Rajcevic, B., Guimarães, N.C.C., van Droogenbroeck, B., Valicente, F.H. *et al.* (2012). Detection of airborne genetically modified maize pollen by real-time PCR. *Mol. Ecol. Resour.*, 12, 810–821.

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. *Mol. Mar. Biol. Biotech.*, 3, 294–299.

Fonseca, V.G., Carvalho, G.R., Sung, W., Johnson, H.F., Power, D.M., Neill, S.P. *et al.* (2010). Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nat. Commun.*, 1, 98.

Gardner, T.A., Burgess, N.D., Aguilar-Amuchastegui, N., Barlow, J., Berenguer, E., Clements, T. *et al.* (2012). A framework for integrating biodiversity concerns into national REDD+ programmes. *Biol. Conserv.*, 154, 61–71.

Ghodsi, M., Liu, B. & Pop, M. (2011). DNACLUST: accurate and efficient clustering of phylogenetic marker genes. *BMC Bioinformatics*, 12, 271.

Gibbons, J.M., Nicholson, E., Milner-Gulland, E.J. & Jones, J.P.G. (2011). Should payments for biodiversity conservation be based on action or results? *J. Appl. Ecol.*, 48, 1218–1226.

Glenn, T.C. (2011). Field guide to next-generation DNA sequencers. *Mol. Ecol. Resour.*, 11, 759–769.

Goldberg, C.S., Pilliod, D.S., Arkle, R.S. & Waits, L.P. (2011). Molecular detection of vertebrates in stream water: a demonstration using rocky mountain tailed frogs and Idaho giant salamanders. *PLoS ONE*, 6, e22746.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A.C. & Baird, D.J. (2011). Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE*, 6, e17497.

Hajibabaei, M., Spall, J.L., Shokralla, S. & van Konynenburg, S. (2012). Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecol.*, 12, 28.

Hao, X., Jiang, R. & Chen, T. (2011). Clustering 16S rRNA for OTU prediction: a method of unsupervised Bayesian clustering. *Bioinformatics*, 27, 611–618.

Hiiesalu, I., Öpik, M., Metsis, M., Lilje, L., Davison, J., Vasar, M. *et al.* (2012). Plant species richness belowground: higher richness and new patterns revealed by next-generation sequencing. *Mol. Ecol.*, 21, 2004–2016.

Janzen, D.H., Hajibabaei, M., Burns, J.M., Hallwachs, W., Remigio, E. & Hebert, P.D.N. (2005). Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 360, 1835–1845.

Jerde, C.L., Mahon, A.R., Chadderton, W.L. & Lodge, D.M. (2011). "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conserv. Lett.*, 4, 150–157.

Kermarrec, L., Franc, A., Rimet, F., Chaumeil, P., Humbert, J.F. & Bouchez, A. (2013). Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms. *Mol. Ecol. Resour.*, 13, 607–619.

Kinzig, A.P., Perrings, C., Fernández, D.F.D.z., Polasky, S., Smith, V.K., Tilman, D. & Turner, B.L. (2011) Paying for ecosystem services–promise and peril. *Science*, 334, 603–604.

Kircher, M. & Kelso, J. (2010). High-throughput DNA sequencing - concepts and limitations. *BioEssays*, 32, 524–536.

Kleijn, D., Rundlöf, M., Scheper, J., Smith, H.G. & Tscharntke, T. (2011). Does conservation on farmland contribute to halting the biodiversity decline? *Trends Ecol. Evol.*, 26, 474–481.

Knight, A.T., Bode, M., Fuller, R.A., Grantham, H.S., Possingham, H.P., Watson, J.E.M. *et al.* (2010). Barometer of life: more action, not more data. *Science*, 329, 141.

Lenz, T. & Becker, S. (2008). Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci — Implications for evolutionary analysis. *Gene*, 427, 117–123.

Levi, T., Shepard, G.H. Jr, Ohl-Schacherer, J., Peres, C.A. & Yu, D.W. (2009). Modelling the long-term sustainability of indigenous hunting in Manu National Park, Peru: landscape-scale management implications for Amazonia. *J. Appl. Ecol.*, 46, 804–814.

Lindenmayer, D. & Likens, G. (2011). Direct measurement versus surrogate indicator species for evaluating environmental change and biodiversity loss. *Ecosystems*, 14, 47–59.

Matsen, F.A., Kodner, R.B. & Armbrust, E.V. (2010). pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, 11, 538.

Meijaard, E. & Sheil, D. (2012). The dilemma of green business in tropical forests: how to protect what it cannot identify. *Conserv. Lett.*, 5, 342–348.

Munch, K., Boomsma, W., Huelsenbeck, J., Willerslev, E. & Nielsen, R. (2008). Statistical assignment of DNA sequences using Bayesian phylogenetics. *Syst. Biol.*, 57, 750–757.

Newton, A.C. (2011). Implications of Goodhart's Law for monitoring global biodiversity loss. *Conserv. Lett.*, 4, 264–268.

Nicholson, E., Collen, B., Barausse, A., Blanchard, J.L., Costelloe, B.T., Sullivan, K.M.E. *et al.* (2012). Making robust policy decisions using global biodiversity indicators. *PLoS ONE*, 7, e41128.

Nolte, V., Pandey, R.V., Jost, S., Medinger, R., Ottenwälder, B., Boenigk, J. *et al.* (2010). Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Mol. Ecol.*, 19, 2908–2915.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, R.B., O'Hara, R.B. *et al.* (2012). *vegan*: Community Ecology Package. R Package version 2.0-5. Available at: http://CRAN.R-project.org/package=vegan.

Pedley, S.M., Franco, A.M.A., Pankhurst, T. & Dolman, P.M. (2013). Physical disturbance enhances ecological networks for heathland biota: a multiple taxa experiment. *Biol. Conserv.*, 160, 173–182.

Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J. *et al.* (2013). Essential biodiversity variables. *Science*, 339, 277–278.

Pompanon, F., Deagle, B.E., Symondson, W.O.C., Brown, D.S., Jarman, S.N. & Taberlet, P. (2011). Who is eating what: diet assessment using next generation sequencing. *Mol. Ecol.*, 21, 1931–1950.

Porazinska, D.L., Giblin-Davis, R.M., Faller, L., Farmerie, W., Kanzaki, N., Morris, K. *et al.* (2009). Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Mol. Ecol. Resour.*, 9, 1439–1450.

Porazinska, D.L., Giblin-Davis, R.M., Esquivel, A., Powers, T.O., Sung, W. & Thomas, W.K. (2010). Ecometagenetics confirms high tropical rainforest nematode diversity. *Mol. Ecol.*, 19, 5521–5530.

Possingham, H.P., Wintle, B.A., Fuller, R.A. & Joseph, L.N. (2012). The conservation return on investment from ecological monitoring. In *Biodiversity Monitoring in Australia*. (eds Lindenmayer, D.B., Gibbon, P.). CISRO Publishing, Australia, pp. 49–58.

R Core Team (2012). *R: A language and environment for statistical computing*. In, R Foundation for Statistical Computing Vienna, Austria.

Ranwez, V., Harispe, S., Delsuc, F. & Douzery, E.J.P. (2011). MACSE: multiple alignment of coding sequences accounting for frameshifts and stop codons. *PLoS ONE*, 6, e22594.

Ratnasingham, S. & Hebert, P.D.N. (2007). BOLD: the barcode of life data system. *Mol. Ecol. Notes*, 7, 355–364.

Rougerie, R., Smith, M.A., Fernandez-Triana, J., Lopez-Vaamonde, C., Ratnasingham, S. & Hebert, P.D.N. (2010). Molecular analysis of parasitoid linkages (MAPL): gut contents of adult parasitoid wasps reveal larval host. *Mol. Ecol.*, 20, 179–186.

Schnell, I.B., Thomsen, P.F., Wilkinson, N., Rasmussen, M., Jensen, L.R.D., Willerslev, E. *et al.* (2012). Screening mammal biodiversity using DNA from leeches. *Curr. Biol.*, 22, R262–R263.

Stuart, S.N., Wilson, E.O., McNeely, J.A., Mittermeier, R.A. & Rodriguez, J.P. (2010). The barometer of life. *Science*, 328, 177–177.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.*, 21, 2045–2050.

Takahara, T., Minamoto, T. & Doi, H. (2013). Using environmental DNA to estimate the distribution of an invasive fish species in ponds. *PLoS ONE*, 8, e56584.

Thomsen, P.F., Kielgast, J., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T.P. *et al.* (2011). Monitoring endangered freshwater biodiversity using environmental DNA. *Mol. Ecol.*, 21, 2565–2573.

Thomsen, P.F., Kielgast, J., Iversen, L.L., Møller, P.R., Rasmussen, M. & Willerslev, E. (2012). Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS ONE*, 7, e41732.

Warton, D.I., Wright, S.T. & Wang, Y. (2012). Distance-based multivariate analyses confound location and dispersion effects. *Methods Ecol. Evol.*, 3, 89–101.

Wiens, J., Sutter, R., Anderson, M., Blanchard, J., Barnett, A., Aguilar-Amuchastegui, N. *et al.* (2009). Selecting and conserving lands for biodiversity: the role of remote sensing. *Remote Sens. Environ.*, 113, 1370–1381.

Wintle, B.A., Runge, M.C. & Bekessy, S.A. (2010). Allocating monitoring effort in the face of unknown unknowns. *Ecol. Lett.*, 13, 1325–1337.

Wunder, S. (2008). Necessary conditions for ecosystem service payments. In: *Economics and conservation in the tropics: a strategic dialogue*. Conservation Strategy Fund Washington DC, USA, pp. 1–11.

Yoccoz, N.G., Bråthen, K.A., Gielly, L., Haile, J., Edwards, M.E., Goslar, T. *et al.* (2012). DNA from soil mirrors plant taxonomic and growth form diversity. *Mol. Ecol.*, 21, 3647–3655.

Yu, D.W. (2010). Managing the exploitation of wildlife in tropical forests. In *Conservation Biology for All*. (eds Sodhi, N.S., Ehrlich, P.R.). Oxford University Press, Oxford, pp. 121–123.

Yu, D.W., Ji, Y.Q., Emerson, B.C., Wang, X.Y., Ye, C.X., Yang, C.Y. *et al.* (2012). Biodiversity Soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods Ecol. Evol.*, 3, 613–623.

Zabel, A. & Roe, B. (2009). Optimal design of pro-conservation incentives. *Ecol. Econ.*, 69, 126–134.

Zhan, A., Hulák, M., Sylvester, F., Huang, X., Adebayo, A.A., Abbott, C.L. *et al.* (2013). High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. *Methods Ecol. Evol.*, 4, 558–565.

Zhang, A.B., Muster, C., Liang, H.B., Zhu, C.D., Crozier, R., Wan, P. *et al.* (2012). A fuzzy-set-theory-based approach to analyse species membership in DNA barcoding. *Mol. Ecol.*, 21, 1848–1863.

Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L. *et al.* (2013). Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, 2, 4.

SUPPORTING INFORMATION

Additional Supporting Information may be downloaded via the online version of this article at Wiley Online Library (www.ecologyletters.com).

Editor, Marcel Holyoak
Manuscript received 6 March 2013
First decision made 11 April 2013
Second decision made 23 June 2013
Manuscript accepted 1 July 2013