## **Methods in Ecology and Evolution**

Methods in Ecology and Evolution 2013, 4, 383-392

doi: 10.1111/2041-210X.12031

# Comparing diversity data collected using a protocol designed for volunteers with results from a professional alternative

### Ben G. Holt<sup>1,2\*</sup>†, Rodolfo Rioja-Nieto<sup>2,3</sup>, M. Aaron MacNeil<sup>4</sup>, Jan Lupton<sup>2</sup> and Carsten Rahbek<sup>1</sup>

<sup>1</sup>Department of Biology, Center for Macroecology, Evolution and Climate, University of Copenhagen, Universitetsparken 15, Copenhagen DK-2100, Denmark; <sup>2</sup>Centre for Marine Resource Studies, School for Field Studies, 1 West Street, South Caicos, Turks and Caicos Islands, British West Indies; <sup>3</sup>Facultad de Ciencias, Unidad Multidisciplinaria de Docencia e Investigación – Sisal, Universidad Nacional Autónoma de México, Puerto de Abrigo s/n Hunucmá, Sisal 97355, México; and <sup>4</sup>Australian Institute of Marine Science, PMB 3 Townsville MC, Townsville, Qld 4810, Australia

#### Summary

1. In light of the continuing biodiversity crisis, the need for high-resolution, broad-scale ecological data is particularly acute. The expansive scale of volunteer data collection programmes provides an opportunity to address this challenge, however, protocols used to collect such data are typically less standardized than those used by professional scientists. Although previous studies have established that different protocols can lead to different results, it remains unclear how relevant these differences are to specific study goals, such as biodiversity assessment.

**2.** This study uses both null model and Bayesian occupancy approaches to examine the capacity of a widely used volunteer survey protocol, the roving diver transect, to detect patterns of marine fish diversity. Richness estimates are compared with those obtained using the conventional belt transects favoured in many peer reviewed studies, examining the power of both protocols to detect statistically significant differences between survey sites and quantifying differences in detectability.

3. Pairwise site comparisons of  $\alpha$ -diversity (i.e. within site diversity) were consistent between protocols, particularly for species totals.

**4.** The roving diver transect protocol detected a substantially larger number of species than the belt transect protocol, due to notably higher detectability, even after controlling for confounding factors. Both protocols detected the same species pool, although the species richness among observations was higher for the belt protocol at certain sites.

5. The significance of pairwise site  $\beta$ -diversity (i.e. differentiation between sites) comparisons differed between the protocols and care should be exercised, when using either protocol, when studying variation in species composition.

**6.** These results provide vital information for managers and researchers considering the use of volunteer data or protocols for the purpose of biodiversity assessment in aquatic systems, helping to quantify the value of thousands of existing survey records. The larger number of species detected by the volunteer protocol suggests this protocol may be advantageous with regards to the completion of taxonomic lists.

**Key-words:** conservation, coral reef, null models, Reef Environmental Education Foundation, roving diver transect, SCUBA, species density, underwater visual census

#### Introduction

One of the most important challenges for biologists is to describe and explain geographical patterns in biodiversity. Analysis of such patterns provides insight into the ecological and evolutionary processes that shape life on earth, and is also a prerequisite for conservation prioritization. Research into

\*Correspondence author. E-mail: bholt@bio.ku.dk

large-scale diversity patterns has traditionally focused on taxonomic groups for which large amounts of distributional data are available, such as birds, mammals and butterflies, with many of these systems benefitting from volunteer data collection (e.g. Robbins *et al.* 1989; Greatorex-Davies & Roy 2000; Newman, Buesching & Macdonald 2003). The value of volunteer schemes with regards to biodiversity monitoring has been considered for these systems, with mixed results (Lovell *et al.* 2009; Schmeller *et al.* 2009; Kremen, Ullmann & Thorp 2011).

There is an urgent need to expand the taxonomic, temporal and spatial scale of applied and theoretical biodiversity

<sup>†</sup>Present address: School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK.

<sup>© 2013</sup> The Authors. Methods in Ecology and Evolution © 2013 British Ecological Society

research, particularly within less accessible environments such as aquatic systems. Paradoxically, some volunteer data collection schemes have been highly successful in these environments with regards to the quantity of data collected (e.g. Pattengill-Semmens & Semmens 2003; Goffredo *et al.* 2010). If these data are shown to be suitable for the study of patterns of diversity, the value of such schemes will hugely increase, with implications for the collection of data in all ecosystems.

A key aspect relating to the value of volunteer data is the reliability of data returned from the protocols used to collect it. For studies performed by professional scientists, underwater visual survey protocols are often designed to minimize bias, maximize precision and ensure repeatability. Due to logistical limitations, vast sections of the world's aquatic ecosystems are rarely, or never, surveyed by professional scientists. The large pool of volunteer enthusiasts has potential to substantially augment the census capabilities of professional researchers. For example, over 8,000 surveys were performed worldwide during 2011 alone by one volunteer organization (R.E.E.F. 2012). Protocols designed for volunteers also attempt to standardize survey efforts, but must balance this requirement against the need to maintain the interest of the public. Whether data produced by such protocols are suitable for comparative studies of biological diversity remains unclear.

The development and popularity of underwater visual survey techniques using self-contained underwater breathing apparatus (SCUBA) equipment has resulted in monitoring of the underwater environment on a scale that was previously impossible. Underwater visual survey methods have been used extensively in tropical (e.g. Pattengill-Semmens & Semmens 2003) and temperate marine habitats (e.g. Goffredo et al. 2010), as well as freshwater systems (e.g. Brosse et al. 2001). Many previous studies have compared underwater visual survey protocols; most of these studies focused on identifying sources of bias within methods, often with a view to quantifying differences among protocol (e.g. Thresher & Gunn 1986; St John, Russ & Gladstone 1990; Sullivan & Chiappone 1992; Miller & Ambrose 2000; Schmitt, Sluka & Sullivan-Sealey 2002). Few, if any, of these studies have been focused on the capacity for underwater visual survey protocols to reflect actual biological patterns or to test specific ecological hypotheses. This is surprising, as it is widely acknowledged that decisions regarding the choice of methodology should be based on the study question. The likely reason for this discrepancy is that in many underwater ecosystems it is impossible to completely sample any area using any method and, without a full taxonomic list for comparison, it is difficult to quantify the performance of any particular sampling method. Our study addresses this issue by concentrating large amounts of survey effort on a very small number of sites (three) to both reliably identify any differences between two test protocols and thoroughly elucidate patterns among study sites. The techniques chosen for this study represent the most frequently used underwater visual survey methodology in published peer reviewed fish diversity studies (the belt transect) and the Roving Diver Technique (RDT) used by the Reef Environmental Education

Foundation (REEF) volunteer fish survey project (Pattengill-Semmens & Semmens 2003), thought to be the largest marine species sighting database in the world, and similar to protocols used by other successful programmes. Volunteer data, such as those collected by REEF, are potentially a highly valuable resource for the marine environment, where the measurement of fundamental aspects of diversity, across expansive spatial scales, has been suggested to be a key management priority (Palumbi et al. 2008). Although studies typically vary considerably on the specific aspects of diversity they address, e.g. taxonomical relatedness (Carranza, Defeo & Arim 2011), phylogenetic diversity, functional diversity (Halpern & Floeter 2008), species diversity and community composition comparisons (i.e.  $\alpha$  and  $\beta$ -diversity) are relevant to most studies and conservation objectives, and are therefore the focus of this study. As belt transects are regularly used in professional reef fish diversity studies (Kulbicki et al. 2010), they represent a logical choice with which to compare the performance of the RDT protocol. The extent to which belt transect results are consistent to those produced using RDT protocols is therefore informative regarding the utility of vast amounts of volunteer data that are currently available and collected in the future. The objective of this study is to determine whether the two protocols differ in terms of the  $\alpha$  (i.e. within site diversity) and β-diversity (i.e. differentiation between sites) of the communities they record and in their power to detect significant differences in these biodiversity measures between these communities. We also examine how detectability (i.e. probability to detect a species that is present in a surveyed area at the time of survey) varies between protocols, as well variation associated with sites, functional groups, taxonomic groups, survey duration and underwater visibility.

#### Materials and methods

#### STUDY DESIGN

The study included a total of 144 underwater visual surveys focused on three sites, with a survey site defined by the precise location at which divers entered the water. All sites were close to Long Cay off South Caicos in the Turks & Caicos Islands (Fig. 1). The survey sites were chosen to represent habitats that might be expected to differ in fish diversity. Our study was conducted at sites that appeared to differ in terms of species richness; based on preliminary visual inspection rather than existing survey data to avoid any bias based on similarity of either of our test protocols to protocols used to collect pre-existing data. Site A comprised of primarily bare rock substrata, with very little benthic biota, and was proposed to have low diversity. Site B primarily comprised of sand with abundant soft corals and very low hard coral cover, and was proposed to have intermediate diversity. Site C represented a fairly healthy coral reef site, with relatively high hard coral cover, and was proposed to have high diversity.

#### SURVEY METHODOLOGY

Surveys were completed by two teams of 12 divers over two periods of 2 weeks during the spring and autumn of 2009, with each team respon-



Fig. 1. Locations of survey sites used for study of belt transect and roving diver transect underwater visual survey protocols. Sites chosen on the basis of expected variation in fish species diversity; site A = low diversity, site B = intermediate diversity and site C = high diversity.

sible for one study period. During each study period, all 12 divers surveyed each of the three sites twice (once using the belt transect protocol, once using the RDT protocol), with the order of surveys alternated among the sites and the protocols used, to address any possible temporal bias in data collection. Sampling effort was identical for each of the two study periods and all data were pooled together (analysis of seasonal trends in these fish communities is not within the scope of this study). For both of the survey protocols tested, dive teams were divided into buddy pairs, with one buddy pair responsible for one survey. Prior to the beginning of the study, all divers completed an intensive fish identification course, which covered over 130 species commonly occurring in the local area. It was rare for surveyors to encounter a species they could not positively identify, and on these occasions, divers took detailed notes on these fish and identified them after returning from the survey trip.

#### PROTOCOLS TESTED

#### Belt transect

At each site, a pair of divers conducted three 50 m long transects that were set approximately 50 m apart, parallel to the isobaths. For each transect, divers positioned themselves 2.5 m either side of and 2.5 m above a transect line and recorded all fish found within the 5 m wide belt transect. Once the transect line was laid out, divers waited for 1 min to allow the fish to settle before beginning the transect. Divers swam along the transect at a rate of 10 m per minute, therefore taking 5 min to complete each transect. For each species, the total number of individuals seen at each transect was recorded. Data from all three transects completed during one dive were pooled.

#### Roving diver transect

During these surveys, divers swam throughout a dive site for a period of approximately 45 min and recorded every fish species seen that could be positively identified. The search for fishes began as soon as the diver entered the water. Divers were encouraged to look under ledges and up in the water column. Each recorded species was assigned one of four abundance categories based on how many were seen throughout the dive [single (1); few (2–10); many (11–100) and abundant (> 100)]. For this study, sighting records were used only as presence/absence data, as no diversity metrics are currently available to include such abundance categories. In addition to fish species observations, divers also reported the time, date, bottom time, visibility, average depth, current strength and habitat category for each dive, in accordance with the REEF volunteer fish survey requirements. All RDT survey data were entered into the REEF volunteer survey project database (www.reef. org/programs/volunteersurvey).

#### DATA ANALYSES

Analyses of species totals and species richness differences were based on null model comparisons and performed in R statistical software (R Development Core Team 2012) and utilizing the vegan package (Oksanen *et al.* 2012). All codes used are available in Supporting Information 1. Sites and protocols are compared in a pairwise manner with empirical data arranged as either a presence/ absence matrix (for RDT and mixed protocol matrices) and abundance matrix (belt transect only matrices), with species as rows and individual surveys as columns. All possible pairwise comparisons were made between sites and between the two protocols, for both  $\alpha$  and  $\beta$  diversity, as follows:

#### α-diversity

Species richness is the most commonly used measure of  $\alpha$ -diversity when sampling communities and reliable abundance data are not available (e.g. Stanley 2007). The precise meaning of species richness varies between studies, but here the definition of Gotelli & Colwell (2001) is used, i.e. the diversity of species per standardized number of individuals. To compare species richness values across samples with differing numbers of individuals, the number of observed species should be 'rarefied' to a consistent number of individuals (Gotelli & Colwell 2001). Species richness is also often used to refer to the number of species within a given area and this diversity measure is often the focus of conservation management. Therefore, this  $\alpha$ -diversity measure was also considered (referred to throughout as species totals).

#### Species totals analysis

Pairwise differences in observed species totals, between sites or between protocols, were tested against a null expectation that both sets of results represent samples of a common species pool, which do not differ in species totals. To simulate a null expectation of equal species totals the empirical matrices were randomized 10 000 times, with the condition that species occurrences/abundances (i.e. row totals) remained fixed to that of the original matrix. Null pairwise differences are calculated using species totals derived from null matrices, after these matrices have been split on the same basis as the empirical data (e.g. Site A species total is based on the original columns for Site A surveys). Empirical differences in species totals were tested against null differences in a 'one-tailed' manner, i.e. proportion of null differences higher than (or equal to) the empirical difference, as it is not possible for empirical differences to be lower than the null expectation of zero.

#### Species richness analysis

As the RDT protocol does not reliably record abundance of individuals, species richness per given number of observations was used as proxy for species richness per given number of individuals (for this protocol only), with an observation being defined as one record of species recorded during a single survey. Pairwise differences in species richness, between sites or between protocols, were tested against the null expectation that both sets of results represent samples from the same species pool, but communities differ in species totals in accordance with the empirical data. To simulate this null expectation, empirical matrices were randomized 10,000 times according to the 'quasiswap' method (Miklós & Podani 2004) for presence/absence matrices and according to Patefield's (1981) algorithm for abundance data, which fixes row and column totals to that of the original matrix. Therefore, both the observed differences in number of individuals/observations for each species and the number of individuals/observations for each survey were maintained during this randomization. Empirical differences in species totals were tested against null differences in a 'two-tailed' manner, i.e. proportion of null differences higher than (or equal to) the empirical difference and proportion of null differences lower than (or equal to) the empirical difference. The lower of the two results is reported as the P value (after multiplying it by two to allow for the multiple testing).

#### $\beta$ -diversity

 $\beta$ -diversity was quantified using Whittaker's (1960) original  $\beta$ -diversity metric, computed as:

$$\beta_{\rm W} = \frac{a+b+c}{(a+b+2c)/2} - 1$$

where a = number of species in one data set (i.e. within one site or one protocol), b = number of species in the other data set (i.e. within other site/protocol) and c = the number of species shared between data sets (i.e. found in both sites or using both protocols). The significance of  $\beta_W$  was tested against the same null model as described for species richness difference tests above. As with the species richness difference tests, empirical differences in species totals were tested against null differences in a 'two-tailed' manner.

#### STATISTICAL POWER OF PROTOCOL DATA

Number of surveys required to detect significant differences between sites was used as a measure of statistical power for each survey protocol, for each pairwise site comparison. At each sampling level (i.e. from 1 to 24 surveys completed), the significance of species totals and species richness differences was tested as explained above. For all sampling levels, except 24, the surveys were randomly selected from those available. This process was repeated 100 times and the median values across all runs calculated. The lowest number of surveys that, on average, detected a significant difference between a pair of sites (median P < 0.05), without higher sampling levels showing a conflicting result, was considered to be the minimum requirement to detect a significant difference between these sites.

#### OCCUPANCY MODELS

Occupancy models have seen widespread application in terrestrial systems and have been used to investigate species diversity patterns, as well species detectability (Kery & Royle 2008), however, these methods have only recently been applied in coral reef studies (Cheal *et al.* 2012). To examine the drivers of detection variation across our samples, we adopted a Bayesian occupancy approach that effectively models the effects of site, species group and sampling characteristics on observed samples.

First, we compiled a list of 295 candidate species that were judged to have a nonzero probability of being present at any of our study locations; these included diurnally active, reef or shallow-flat water associated species whose range overlapped the study sites according to FishBase (fishbase.org). We excluded sharks as they are not generally collected by conventional underwater visual survey methods and require exceptionally large transects for unbiased sampling. For each of the candidate species, we assigned a functional group, taxonomic group (order), maximum total length (TL) and trophic position (TP) as these attributes have been shown to affect their detectability underwater (MacNeil *et al.* 2008a,b).

Secondly, we developed a set of candidate occupancy models that included covariates for both occupancy and detection at each location in the spring or autumn (j = 1, 2, ..., 6). These models quantify the probability of detection for species i at j ( $\theta_{ij}$ ) across K = 24 sampling occasions (i.e. 24 surveys per site per season), conditional on a model-estimated (latent) state of occupancy for i at j ( $z_{ij}$ ):

$$p(z_{ij}|\psi_{ij}) = \psi_{ij}^{z_{ij}} (1 - \psi_{ij})^{1 - z_{ij}}$$
eqn 1

$$p(y_{ij}|\theta_{ij}, z_{ij}) = \binom{K}{Y_{ij}} \theta_{ij}^{y_{ij}} (1 - \theta_{ij})^{K - y_{ij}^{z_{ij}}}$$
eqn 2

Because the latent occupancy state is partially observed (we know  $z_{ij} = 1$  for species we have observed at a given site), we can use the detection history for species that have been observed to estimate the probability of detection for species not observed and, by consequence, estimate the probability of their presence given nondetection  $(1-\theta_{ij})$ . This is the approach taken by Dorazio & Royle (2005) and Dorazio *et al.* (2006), albeit with slightly different implementation to our own. Probabilities of detection and occupancy were modelled simultaneously, using covariates relevant to each part:

 $logit(\theta_{ii}) = X_d \gamma$  eqn 3

 $logit(\psi_{ij}) = X_o \beta$  eqn 4

where  $\gamma$  and  $\beta$  are vectors of normally distributed parameters, given the covariate matrix for detection (X<sub>d</sub>) and occupancy (X<sub>o</sub>). Potential covariates for detection included species, site, and survey-specific attributes (Table 1), in particular the use of RDT or belt transect methods. All models were run using the PyMC package (Patil, Huard & Fonnesbeck 2010) for the Python programming language (http://www.python.org). Hyperparameters for the Normal means and precisions present in models (3) and (4) were given weakly informative priors [means approximately N(0, 0.01); precisions approximately  $U(0,1000)^{-2}$ ]. Models were run for 40 000 iterations, after a 60 000 iteration burn-in period, and examined for autocorrelation and convergence using the posterior plotting features provided by PyMC. Model selection was conducted using the deviance information criterion (DIC) and fit was examined through Bayesian P values that compare expected vs. observed model deviances, with values < 0.05 or > 0.95 being evidence for substantial lack of fit (Gelman et al. 2004). Both the data and sample PyMC code are provided in supporting information 2-9, and in a GitHub repository (https://github.com/mamacneil/volunteer uvc).

#### Results

Full survey details and data are available in Supporting Information 10. A total of 140 species were recorded across all surveys, with 105 species detected at the site A, 114 at site B and 119 at site C.

#### GENERAL COMPARISON OF PROTOCOLS

Across all three sites, the total number of species detected after all surveys was higher using the RDT protocol (137) than for the belt transect protocol (106), with this difference significantly larger than expected under the equal species totals null model (difference = 31, mean null difference = 4.049, P < 0.001, Fig. 2a). This difference did not exceed expectations under the equal species richness null model (mean null difference = 34.87, P = 0.789, Fig. 2a). The  $\beta_W$  value produced for the comparison of the two protocols was not significantly different to null expectations ( $B_W = 0.152$ , mean null  $B_W = 0.169$ , P = 0.480, Fig. 2b). Within sites, species totals and  $\beta_W$  results were consistent to that seen when pooling sites, i.e. differences in species totals between protocols were significantly higher than expected, but  $\beta_W$  fell within null expectations. However, within-site species richness results were not consistent, such that sites A and B showed significantly lower differences in species richness between protocols than null expectations (Fig. 2a).

#### PAIRWISE SITE COMPARISONS

#### Species totals

Both protocols ranked the three survey sites in the same order in terms of species totals (Fig. 3a). The general significance of pairwise species totals differences between sites according to the RDT data (differences = 12, 19, 7; mean null differences = 4.45, 4.24, 4.12; P = 0.022, 0.001, 0.146; for A vs. B, A vs. C and B vs. C respectively) was consistent with that seen by the belt transect data (differences = 12, 16, 4; mean null differences = 4.58, 4.30, 4.32; P = 0.025, 0.002, 0.407; for A vs. B, A vs. C and B vs. C respectively), with sites B and C being significantly higher in species totals than site A but not significantly different from each other (Fig. 4).

 Table 1. Covariate and deviance information criterion (DIC) values for reef fish occupancy models across three sites in two seasons in Turks and Caicos. Covariates include method Roving Diver Technique (RDT or belt transect), FG (functional group), taxonomic order, duration (min) and visibility (m). Location indicates the combination of site and season over which fish communities are assumed closed

Model	Detection	Occupancy	DIC	ΔDIC
M0	Intercept	Intercept	17439	1421
M1	Intercept + Method	Intercept	16609	591
M2	Intercept + Method	Location*	16609	591
M3	Method + $FG^*$	Location*	16044	26
M4	Method $+$ FG* $+$ Duration $+$ Visibility	Intercept	16031	13
M5	Method $+$ FG* $+$ Duration $+$ Visibility	Location*	16018	0
M6	Method + $FG^*$ + Duration + Visibility + $TL$ + $TP$	Location*	16042	24
M7	Method $+$ FG* $+$ Duration $+$ Visibility	Location* + Order*	NA	NA
M8	Method $+$ FG* $+$ Duration $+$ Visibility	Order*	16048	30
M9	Method + $FG^*$ + Duration + Visibility	Location* + FG*	NA	NA
M10	Method $+$ FG* $+$ Duration $+$ Visibility	FG*	17185	1167
M11	Method + FG* + Duration + Visibility + Site*	Location*	NA	NA
M12	Method + FG* + Duration + Visibility + Site*	Site* + season	1636	18
M13	Method + FG* + Duration + Visibility + Site* + Observer team +	Location*	1624	6

Models marked \* are hierarchical; NA's indicate models that failed to converge.



Fig. 2. (a) Difference in numbers of species recorded between belt transect and RDT SCUBA survey protocols. Black dots represent observed differences across and within survey sites. Crosses represent one-tailed 95% confidence level for null expectation of equal species totals between protocols. Error bars represent two-tailed 95% confidence level for null expectation of equal species totals between protocols, based on 10 000 random simulations. (b)  $\beta$ -diversity comparisons for belt transect and RDT SCUBA survey protocols, calculated as Whittaker's  $\beta$ -diversity metric ( $\beta_W$ ). Black dots represent observed  $\beta_W$  across and within survey sites. Error bars represent two-tailed 95% confidence level for null expectation of equal species totals between protocols, eaculated as Whittaker's  $\beta$ -diversity metric ( $\beta_W$ ). Black dots represent observed  $\beta_W$  across and within survey sites. Error bars represent two-tailed 95% confidence level for null expectation of equal species richness and unequal species totals between protocols, based on 10 000 random simulations.



Fig. 3. Species accumulation curves with increasing survey effort based on random subsamples (surveys sampled without replacement) of the overall data set (10 000 resamples), for RDT (solid lines) and belt transects (dashed lines), at sites A (red), B (green) and C (blue). See main text for site descriptions. Shaded areas represent 95% boundaries of observed totals (N.B., these are constrained by the number of surveys completed and are not suitable for hypothesis testing). (a) Site species total accumulation curves, all species totals scaled by number of surveys completed. (b) Species richness accumulation curves, belt transect species totals scaled by number of individuals detected, RDT species totals scaled by number of observations made. Numbers of individuals/observations shown as percentage of maximums per site (1307 observations for RDT data, 6056 individuals for belt transects, both at site C) for figure plotting purposes.

Both the RDT protocol and the belt transect protocol tended to show significant species total differences between sites A and B after 24 surveys (Fig. 4a & b). Differences in species totals between sites A and C tended to be significant after one survey for both the RDT and the belt transect protocols (Fig. 4c & d).

#### Species richness

Both RDT and belt transect methods rank the site A lowest in terms of rarefied species richness, however, the results differ between protocols for the ranking of the sites B and C (RDT = 119, 124-1, 130-2 and belt transects = 68, 77-7, 74-3, rarefied species richness for site A, site B and site C respectively, rarefied to 850 observations for RDTs and 3411 individuals for belt transects (Fig. 3)). Neither the RDT data nor the belt transect data showed any significant differences from null species richness model for pairwise comparisons between sites (RDT: differences = 12, 19, 7; mean null differences = 10-78, 24-46, 15-51; P = 0.364, 0.870, 0.955; belt transect: differences = 12, 16, 4; mean null differences = 10-77, 8-67; P = 0.123, 0.794; for A vs. C and B vs. C, respectively, for A vs. B, A vs. C and B vs. C respectively).

#### $\beta_w$ diversity

Overall the survey protocols showed the same  $\beta_w$  relationships among sites, but with belt transect data giving higher values. Sites B and C were more similar to each other (belt  $\beta_w = 0.220$ ; RDT  $B_w = 0.148$ ) than either was to site A (belt  $\beta_w = 0.257$ , 0.263; RDT  $\beta_w = 0.176$ , 0.213; for A vs. B and A vs. C, respectively, Fig. 4).  $\beta_w$  values produced by data from both protocols were consistently higher than mean null expectations; however, the significance of these differences was not consistent between protocols. For the RDT data only the  $\beta_w$  value concerning sites A and C was significantly higher than null expectations (mean null  $\beta_w = 0.157$ , P < 0.001), with the remaining site comparisons returning values that fell within null expectations (mean



Fig. 4. Empirical and null species total differences and  $\beta$ -diversity measures for pairwise site comparisons, based on data collected using belt transect and RDT protocols. Empirical differences in species totals are shown as black dots. 95% percentile for null differences based on equal species totals null model shown as solid line. 95% boundaries for null differences based on unequal species totals (equal species richness) null model shown as grey shading. Empirical Whittaker's  $\beta$ -diversity values ( $\beta_W$ ) shown as crosses, with upper 95% null boundary shown as dotted lines. See main text for site and null model descriptions.

null differences = 0.155, 0.128; P = 0.184, 0.146; for A vs. B and B vs. C respectively). All  $\beta_w$  values for belt transect pairwise site comparisons were significantly higher than null expectations (mean null differences = 0.172, 0.167, 0.152; All P < 0.001; for A vs. B, B vs. C and A vs. C respectively).

Significant  $\beta_w$  values were apparent between site A and site C for the analysis of the RDT data after four surveys (Fig. 4d). For the belt transect results, significant  $\beta_w$  values were apparent after one survey for both comparisons involving site A (Fig. 4a and c) and after two surveys for the site B vs. site C comparison (Fig. 4e).

#### OCCUPANCY MODELLING

The top (DIC-ranked) occupancy model (M5, see Table 1) included site-varying occupancy and detection varying by

functional group, time spent sampling, visibility underwater and the survey method used. There was a considerable difference in detection between methods (Fig. 5a), with average RDT detection [0.38 (0.36,0.40)] being nearly twice that of the belt transect method [0.21 (0.20,0.22)]. In addition, there were positive effects on detection given longer observation periods and increased visibility, and detection varied markedly among functional groups (Fig. 5b), with herbivores being observed most readily ( $\hat{\theta} = 0.59$  [0.44,0.51]; posterior median [95% uncertainty interval]) and the sand-dwelling gobies and blennies being nearly undetectable ( $\hat{\theta} = 0.02$ ) [0.01,0.04]). Estimated species occupancy at each site was somewhat greater under the occupancy model framework than the raw counts at each location (Supporting Information 11), with between 25 and 41 additional species estimated to be present across locations.



Fig. 5. Average detection probabilities per sample by (a) sampling characteristics and (b) functional group, for the top-ranked occupancy model of Turks and Caicos coral reef fish communities. Points are marginal median detection probabilities for each characteristic; lines indicate 95% (thin) and 50% (thick) posterior uncertainty intervals. Dashed vertical lines indicate 50% probability of detection.

#### Discussion

Our results provide evidence that less standardized survey protocols used by volunteer programs may give results that are broadly consistent with those based on methods used by professional scientists. In this study the evaluated survey protocols were highly consistent with regards to comparisons of site species totals. The species richness results show differences in site ranking between the protocols, neither protocol detected any significant pairwise difference between sites. Both protocols show similar  $\beta$ -diversity relationships among sites, but the significance of  $\beta_w$  values was inconsistent between the protocols. After 72 surveys per protocol, RDT surveys record significantly more species than surveys using belt transects, due to substantially higher detectability, i.e. the RDT protocol was capable of recording considerably more species per survey. Despite the observed differences in detectability between the protocols,  $\beta_W$  analysis suggests that there is no significant difference between the protocols regarding the composition of species detected (Fig. 2).

The large number of replicates (n = 24) taken at each location in each season is at the upper end of replication for reef fish surveys, near the point at which most species capable of being detected will be observed, even in low-detection areas (MacNeil et al. 2008b). As completing more than the 24 surveys per site (per protocol) is probably beyond the reach of many survey programmes, RDT may be preferable if the research goal is to detect as many species as possible. Although previous work has suggested that both RDT and belt transects record distinct subsets of the overall species pool (Schmitt, Sluka & Sullivan-Sealey 2002), our analysis suggests that such differences may not differ from null expectations. Protocol species richness comparisons within sites A and B show that the belt transect recorded a higher diversity of species among observations (Fig. 2a), but the RDT protocol more than compensates for this by returning a larger number of observations per survey. The higher species totals of RDT survey data may be driven by factors such as time spent recording fish (Fig. 5a), area covered by the survey and/or flexibility in search methodology. Note that the actual time spent recording fish differs considerably between the two protocols (see Materials and Methods), and, as none of the sites appear to have been sampled to completion

(Fig. 3), this factor may have a strong influence on this result.

Many species will be missed by both sampling protocols, and further work is required to quantify the performance of other types of underwater visual survey protocols, such as stationary counts, against those tested here. It is clear from our occupancy model results that the survey protocols tested are biased towards detecting certain functional groups (Fig. 5a.), ranging from gobies/blennies, which have extremely low detectability, to herbivores, which were observed relatively easily. Destructive methods, such as rotenone sampling, can produce more complete taxonomic inventories than underwater visual survey methods (Smith-Vaniz, Jelks & Rocha 2006); however, although these methods may be suitable for exhaustive sampling, they cannot be applied across extensive spatial scales, or within protected areas, and are therefore limited in terms of their coverage. In addition, these methods also tend to miss some species that are recorded by underwater visual survey methods (Smith-Vaniz, Jelks & Rocha 2006) and such methods may need to be combined if the study goal is to produce a full species inventory.

Species totals may be a more relevant diversity measure for conservation purposes (Gotelli & Colwell 2001) and it is common for studies to be concerned with the number of species in a given area rather than the number of species per given number of individuals (often the term species richness is used to refer to species totals). On the basis of this study, RDTs are recommended as this protocol gave results that were consistent with belt transect data while recording a larger number of species per site. A caveat to this recommendation is that relative area of survey sites should be considered as RDTs are not restricted in terms of the exact area they cover. For this study, all three sites were part of a relatively expansive coastal area and the amount of area covered was determined by dive times and swimming speeds that did not vary substantially between sites. However, if more restricted sites, such as wrecks or small reef systems, are surveyed then differences in habitat area may need to be controlled for. Further studies of this type should support our analytical approach and survey design for additional sites, including other ecosystems, such as terrestrial and freshwater communities.

With regard to species richness, differences seen between protocols at sites A and B are of concern. It is possible that species richness among observations is a poor substitute for species richness among individuals for one or both protocols. Adapting the protocol to include estimates of the exact numbers of individuals for each species may bring benefits in this regard, although any modification would also need to account for variation in detection rate among species (MacNeil *et al.* 2008a,b). Currently, belt transects are preferable because they return specific measures of individual abundance required for this measure of diversity. New analytical approaches also have the potential to address this issue (Yamaura *et al.* 2011).

The inconsistency between the protocols regarding the significance of between-site  $\beta_W$  values is surprising given the protocol comparison results. It is possible that the constrained nature of the null models influences the probability of type II errors for presence/absence data differently to abundance data. The two  $\beta_w$  values that did not differ significantly from null expectations in the RDT results were both close to the upper 95% confidence limit for nearly all sampling levels (Fig. 4). However, these results also suggest that belt transect data return higher ßw values than the RDT data, a result that will be driven by differences in data collection protocols and is not influenced by the existence or absence of abundance data. As ßw values between the two protocols were not significantly different to null expectations, this cannot be explained by the protocols detecting different species. By values tended to decline as more surveys are completed (Fig. 4) and therefore the higher number of species returned by the RDT protocols may have resulted in lower (and possibly more accurate)  $\beta_W$  values.

Generally speaking the RDT protocol is successful in terms of the quantity of data that have been collected. The REEF volunteer fish survey project has collected over six million sightings across over 10 000 locations and a similar program in Italy was highly successful at collecting a very large amount of data in a short period of time (Goffredo *et al.* 2010). The results of this study suggest that RDT protocols can be consistent with belt transects when quantifying  $\alpha$ -diversity, providing an invaluable resource for large-scale ecological studies of bio-diversity.

#### Acknowledgements

We thank John Claydon, Anke Loewa, Jonathan Brown, Jim Catlin and William Maclennan for their invaluable assistance in the field. We particularly thank the 24 student divers for their remarkable efforts collecting the data for this project; (all names are provided in Supporting Information 10). We thank Andrea Baquero for assistance in compiling the functional data analysed in this study. BGH and CR acknowledge the Danish National Research Foundation for support to the Center for Macroecology, Evolution and Climate. BGH also thanks the Marie Curie Actions under the Seventh Framework Programme (PIEF-GA-2009-252888). The authors thank Dr Marc Kéry and two anonymous reviewers for their valuable comments and resulting improvements to this manuscript.

#### References

- Carranza, A., Defeo, O. & Arim, M. (2011) Taxonomic relatedness and spatial structure of a shelf benthic gastropod assemblage. *Diversity and Distributions*, 17, 25–34.
- Cheal, A.J., Emslie, E., MacNeil, M.A., Millar, I. & Sweatman, H. (2012) Spatial variation in the functional characteristics of herbivorous fish communities and the resilience of coral reefs. *Ecological Applications*, Online early http://dx.doi. org/10.1890/11-2253.1 [accessed 14 October 2012].
- Dorazio, R.M. & Royle, A.J. (2005) Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the Ameri*can Statistical Association, **100**, 389–398.
- Dorazio, R.M., Royle, A.J., Söderström, B. & Glimskär, A. (2006) Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology*, 87, 842–854.
- Gelman, A., Carlin, J., Stern, H. & Rubin, D. (2004) Bayesian Data Analysis. Chapman and Hall, New York, USA.
- Goffredo, S., Pensa, F., Neri, P., Orlandi, A., Gagliardi, M.S., Velardi, A., Piccinetti, C. & Zaccanti, F. (2010) Unite research with what citizens do for fun: "recreational monitoring" of marine biodiversity. *Ecological Applications*, 20, 2170–2187.
- Gotelli, N.J. & Colwell, R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, 4, 379–391.
- Greatorex-Davies, J.N. & Roy, D.B. (2000) The butterfly monitoring scheme: report to recorders 1999. Institute of Terrestrial Ecology, http://www.ukbms.org/ [accessed 10 June 2012].
- Halpern, B.S. & Floeter, S.R. (2008) Functional diversity responses to changing species richness in reef fish communities. *Marine Ecology Progress Series*, 364, 147–156.
- Kery, M. & Royle, J.A. (2008) Hierarchical Bayes estimation of species richness and occupancy in spatially replicated surveys. *Journal of Applied Ecology*, 45, 589–598.
- Kremen, C., Ullmann, K.S. & Thorp, R.W. (2011) Evaluating the quality of citizen-scientist data on pollinator communities. *Conservation Biology*, 25, 607– 617.
- Kulbicki, M., Cornuet, N., Vigliola, L., Wantiez, L., Moutham, G. & Chabanet, P. (2010) Counting coral reef fishes: interaction between fish life-history traits and transect design. *Journal of Experimental Marine Biology and Ecology*, 387, 15–23.
- Lovell, S., Hamer, M., Slotow, R. & Herbert, D. (2009) An assessment of the use of volunteers for terrestrial invertebrate biodiversity surveys. *Biodiversity and Conservation*, 18, 3295–3307.
- MacNeil, M.A., Graham, N.A.J., Conroy, M.J., Fonnesbeck, C.J., Polunin, N.V.C., Rushton, S.P., Chabanet, P. & McClanahan, T.R. (2008a) Detection heterogeneity in underwater visual-census data. *Journal of Fish Biology*, 73, 1748–1763.
- MacNeil, M.A., Tyler, E., Fonnesbeck, C., Rushton, S.P. & Conroy, M.J. (2008b) Accounting for detectability in Reef-Fish biodiversity estimates. *Marine Ecology Progress Series*, 367, 249–260.
- Miklós, I. & Podani, J. (2004) Randomization of presence-absence matrices: comments and new algorithms. *Ecology*, 85, 86–92.
- Miller, A.W. & Ambrose, R.F. (2000) Sampling patchy distributions: comparison of sampling designs in rocky intertidal habitats. *Marine Ecology Progress* Series, 196, 1–14.
- Newman, C., Buesching, C.D. & Macdonald, D.W. (2003) Validating mammal monitoring methods and assessing the performance of volunteers in wildlife conservation - "sed quis custodiet ipsos custodies?" *Biological Conservation*, 113, 189–197.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H. & Wagner, H. (2012) Vegan: Community Ecology Package. R Package version 2.0-5. URL http://cran.r-project.org/ web/packages/vegan/.
- Palumbi, S.R., Sandifer, P.A., Allan, J.D., Beck, M.W., Fautin, D.G., Fogarty, M.J., Halpern, B.S., Incze, L.S., Leong, J.-A., Norse, E., Stachowicz, J.J. & Wall, D.H. (2008) Managing for ocean biodiversity to sustain marine ecosystem services. *Frontiers in Ecology and the Environment*, 7, 204–211.
- Patefield, W.M. (1981) Algorithm AS 159: an efficient method of generating random R × C tables with given row and column totals. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 30, 91–97.
- Patil, A., Huard, D. & Fonnesbeck, C.J. (2010) PyMC 2.0: Bayesian stochastic modelling in Python. *Journal of Statistical Software*, 35, 1–81.
- Pattengill-Semmens, C.V. & Semmens, B.X. (2003) Conservation and management applications of the reef volunteer fish monitoring program. *Environmen*tal Monitoring and Assessment, 81, 43–50.
- R Development Core Team (2012) R: A Language and Environment for Statistical Computing. Austria, Vienna.
- Brosse, S., Laffaille, P., Gabas, S. & Lek, S. (2001) Is scuba sampling a relevant method to study fish microhabitat in lakes? Examples and comparisons for three European species. *Ecology of Freshwater Fish*, **10**, 138–146.

#### **392** *B. G. Holt* et al.

- R.E.E.F. (2012) Reef Environmental Education Foundation 2011 Annual report. www.reef.org [accessed 10 June 2012].
- Robbins, C.S., Sauer, J.R., Greenberg, R.S. & Droege, S. (1989) Population declines in North-American birds that migrate to the neotropics. *Proceedings of* the National Academy of Sciences of the United States of America, 86, 7658–7662.
- Schmeller, D.S., Henry, P.Y., Julliard, R., Gruber, B., Clobert, J., Dziock, F., Lengyel, S., Nowicki, P., Deri, E., Budrys, E., Kull, T., Tali, K., Bauch, B., Settele, J., Van Swaay, C., Kobler, A., Babij, V., Papastergiadou, E. & Henle, K. (2009) Advantages of volunteer-based biodiversity monitoring in Europe. *Conservation Biology*, 23, 307–316.
- Schmitt, E.F., Sluka, R.D. & Sullivan-Sealey, K.M. (2002) Evaluating the use of roving diver and transect surveys to assess the coral reef fish assemblage off southeastern Hispaniola. *Coral Reefs*, **21**, 216–223.
- Smith-Vaniz, W.F., Jelks, H.L. & Rocha, L.A. (2006) Relevance of cryptic fishes in biodiversity assessments: a case study at Buck Island Reef National Monument. St. Croix. *Bulletin of Marine Science*, **79**, 17–48.
- St John, J., Russ, G.R. & Gladstone, W. (1990) Accuracy and bias of visual estimates of numbers, size structure and biomass of a coral-reef fish. *Marine Ecology Progress Series*, 64, 253–262.
- Stanley, S.M. (2007) An analysis of the history of marine animal diversity. *Paleobiology*, 33, 1–55.
- Sullivan, K.M. & Chiappone, M. (1992) A comparison of belt quadrant and species presence absence sampling of stony coral (Scleractinia and Milleporina) and sponges for evaluating species patterning on patch reefs of the Central Bahamas. *Bulletin of Marine Science*, 50, 464–488.
- Thresher, R.E. & Gunn, J.S. (1986) Comparative-analysis of visual census techniques for highly mobile, reef-associated piscivores (Carangidae). *Environmental Biology of Fishes*, 17, 93–116.

- Whittaker, R.H. (1960) Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, **30**, 280–338.
- Yamaura, Y., Andrew Royle, J., Kuboi, K., Tada, T., Ikeno, S. & Makino, S. (2011) Modelling community dynamics based on species-level abundance models from detection/nondetection data. *Journal of Applied Ecology*, 48, 67–75.

Received 31 October 2012; accepted 12 January 2013 Handling Editor: Pedro Peres-Neto

#### **Supporting Information**

Additional Supporting Information may be found in the online version of this article.

Data S1. Analytical code of null model biodiversity analysis.

Data S2-S9. Data and sample PyMC code for occupancy models.

Data S10. Survey data collected and analysed for this study.

Data S11. Occupancy-based estimates of species per site.