The iterative process of plant species inventorying for obtaining reliable biodiversity patterns

SILVIA C. ARANDA[1,2]*, HELENA HESPANHOL[3], NÍDIA HOMEM[2,4],
PAULO A. V. BORGES[2,4], JORGE M. LOBO[1,2] and ROSALINA GABRIEL[2,4]

[1]*Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales (CSIC), C/José Gutiérrez Abascal 2, 28006 Madrid, Spain*
[2]*Departamento de Ciências Agrárias, Azorean Biodiversity Group (GBA, CITA-A) and Platform for Enhancing Ecological Research and Sustainability (PEERS), Universidade dos Açores, Rua Capitão João d'Ávila, 9700-042 Angra do Heroísmo, Açores, Portugal*
[3]*CIBIO/InBio, Centro de Investigação em Biodiversidade e Recursos Genéticos da Universidade do Porto, Rua do Campo Alegre, Edifício FC4, s/n, 4169-007 Porto, Portugal*
[4]*CE3C – Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group, Departamento de Ciências Agrárias, Universidade dos Açores, 9700-042 Angra do Heroísmo, Açores, Portugal*

We require representative data of species occurrence to explain plant diversity patterns, but most of the available information is incomplete and biased. To improve our knowledge, we suggest that species inventorying should be an iterative process encompassing the following: (1) the detection of taxonomic and geographical gaps; (2) the planning of a survey design to reduce such gaps; and (3) the evaluation of field sampling results. Here, we focus on the latter phase for the bryophytes of Terceira Island (Azores) for which we have previously estimated < 1% of the area as well surveyed based on historical collections. To examine the performance of our stratified survey based on two factors (land use and environmental regions), we used rarefaction curves, ANOVA tests and bootstrap sampling. We recorded 40% of all the species known for the island and presented eight new citations. The species assemblages remained similar between historical and current inventories. Most localities had completeness values > 85%, but we always exceeded the optimal sampling effort. Land uses and environmental regions affected species diversity, but, unexpectedly, to a different degree. Our study illustrates the difficulties of planning field surveys to obtain reliable biodiversity patterns, even when prior information and standardized sampling protocols are explicitly considered. © 2015 The Linnean Society of London, *Botanical Journal of the Linnean Society*, 2015, **177**, 491–503.

ADDITIONAL KEYWORDS: beta diversity – hornworts – laurel forests – Linnaean shortfall – liverworts – Macaronesia – mosses – spore-producing plants – Wallacean shortfall.

INTRODUCTION

Describing the whole diversity of organisms on the planet is virtually impossible. However, for a certain temporal scale, we can approximate true diversity to numeric estimates that may be more or less reliable, depending on the inventorying effort and the detectability of taxa. For groups such as mammals or birds, our current knowledge of diversity seems to be an accurate estimate based on collection trends and expert knowledge (Chapman, 2009). For hyperdiverse groups, however, there is still much to discover: up to 80% in arthropods (Zhang, 2011; Basset *et al.*, 2012) and probably more than 30% in bryophytes (Bebber *et al.*, 2007; Paton *et al.*, 2008; Mutke & Geffert, 2010), the second most diverse group of plants after angiosperms. These are just examples of the existence

*Corresponding author. E-mail: silviac@mncn.csic.es

of taxonomic gaps, the so-called Linnaean shortfall (Brown & Lomolino, 1998).

In most cases, the Linnaean shortfall is also associated with an unequal knowledge of diversity through geographical space, the so-called Wallacean shortfall (Brown & Lomolino, 1998). That is, we know some areas better than others simply as a result of a matter of spatial scale, opportunity and availability of resources. There are well-known regions for certain taxa, such as British breeding birds (Risely *et al.*, 2011) or water beetles in south-eastern Iberia (Sánchez-Fernández *et al.*, 2008), whereas there are other regions that have been less explored or not explored at all. In Brazil, for example, a country that holds approximately 20% of global angiosperm diversity, there are only 8.7% well-known grid cells at a 0.5° resolution (Sousa-Baena, Garcia & Peterson, 2014). The implications of this bias in the study of diversity patterns are as important as having low values of inventory completeness. If the observed species occurrences do not represent the spatial and environmental spectrum of the study region, it will be difficult to understand the causes of the uneven geographical distribution of species (Sastre & Lobo, 2009; Boakes *et al.*, 2010; Beck *et al.*, 2013).

Despite the long survey tradition carried out over the past two centuries, and our increasing knowledge of the natural history of living organisms, the prevalence of the Linnaean and Wallacean shortfalls stresses the fact that fieldwork is still very necessary for biodiversity studies. Fortunately, there are collection protocols appropriate for each taxon (e.g. Cardoso, 2009; Dengler, 2009; Gaspar *et al.*, 2014) and sampling designs that work well at different spatial scales (e.g. Hirzel & Guisan, 2002; Croft & Chow-Fraser, 2009). For instance, sampling designs based on the environmental diversity approach (Faith & Walker, 1996), such as stratified or heuristic methods, are suitable for maximizing the study of species diversity in a region by accounting for as much environmental heterogeneity as possible (Funk, Richardson & Ferrier, 2005; Medina *et al.*, 2013). Ultimately, however, the degree of inventory completeness of our sampling design will depend on the available resources. The effort that we can invest, both in the field and in the laboratory, is thus the main limiting factor when inventorying biodiversity. In this regard, the existing information on species occurrences, despite being usually incomplete and biased, should play an important role in optimizing the cost-effectiveness of surveys designed to determine the distribution of biodiversity (Sánchez-Fernández *et al.*, 2011).

Furthermore, the available primary biodiversity information may also be highly valuable in validating new sampling data (e.g. Michalcova *et al.*, 2011;

Chytry *et al.*, 2014). We can examine how many species have been collected, and how optimal the invested effort has been, compared with previous knowledge of species diversity. Indeed, after the field and laboratory work, evaluation of the sampling performance is the most important phase in the iterative process of species inventorying (Fig. 1; see also Gaspar *et al.*, 2014). The larger the extent of the study area and the less we know about its diversity, the greater the feedback between the uncertainty present in sampling results and the need for additional data. It is clear from this that the relevance of the past taxonomic and floristic work previously devoted to generate primary biodiversity data is high, despite increasing trends of model-based techniques and massive genetic data to estimate the distribution and diversity of species (Yoccoz, 2012; Beck *et al.*, 2014).
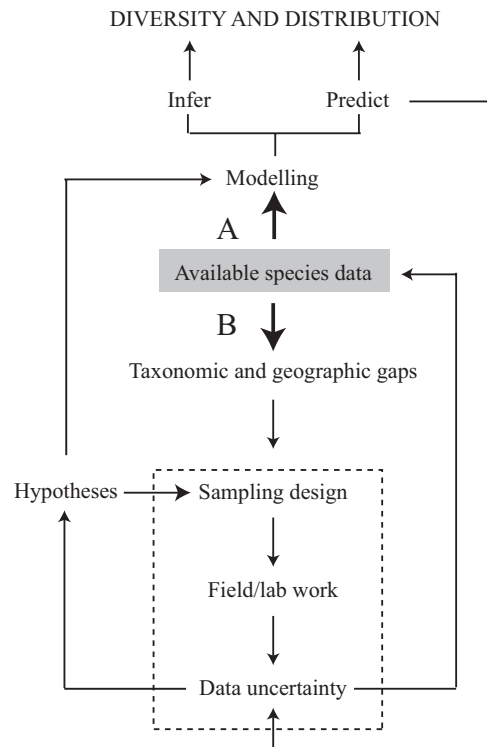

Figure 1. Conceptual framework of the species inventorying process and its implication for ecological and biogeographical studies. If there is sufficient representative information on species occurrences, appropriate modelling techniques can be used to examine our hypotheses on species diversity and distribution (A); otherwise, we need to generate additional data (B). When available, previous information on species occurrences must be considered to optimize and validate the inventorying process (dotted line). Note that the evaluation of sampling results (i.e. inventory completeness) is a key step in the measurement of data uncertainty.

In this study, we exemplify, for the bryophytes of Terceira Island (Azorean archipelago), the strengths and weaknesses of a deliberated field sampling to estimate regional diversity after accounting for all available data previously collected by botanists. Two studies precede this work, which overall illustrate the main sequential steps that should be ideally implemented when inventorying species diversity (Fig. 1). First, we detected important knowledge gaps in this region, despite being the most well-known island in the archipelago with an intensive activity in bryophyte collection during the last 30 years (Aranda *et al.*, 2010). To reduce these gaps, we then planned a stratified sampling covering the environmental and land use variability in the island as a trade-off between the invested working effort and the ecological representation needed to generate reliable diversity estimates (Aranda *et al.*, 2011). Here, we examine the performance of our sampling design by: (1) estimating the inventory completeness and examining its efficiency in terms of sampling effort; (2) comparing the collected diversity values with those provided by historical information; and (3) exploring whether the variables chosen for the survey planning according to the known bryophyte ecology actually explain the diversity patterns obtained. Beyond practical implications, our investigation intends to highlight that species inventorying should be considered as an iterative process that needs to be continuously improved if our aim is to determine which factors and mechanisms are behind biodiversity patterns.

METHODS

AREA OF STUDY

Terceira is a relatively small volcanic island (400 km²) in the central group of the Azorean archipelago, and is separated by approximately 2000 km from mainland Europe, the closest continent. The highest point is Serra S. Bárbara (1021 m). The native vegetation shows temperate oceanic affinities, including the evergreen laurel forests of *Laurus azorica* (Seub.) Franco and *Ilex perado* Aiton ssp. *azorica* Tutin and other vegetation formations dominated by *Juniperus brevifolia* (Seub.) Antoine and several members of Ericaceae. However, these pristine habitats have decreased markedly over time (Gaspar, Borges & Gaston, 2008) and the current landscape of Terceira is a mosaic of cattle pastures and exotic forests of *Cryptomeria japonica* (Thunb. ex L.f.) D.Don, *Eucalyptus* spp. and the invasive species *Pittosporum undulatum* Vent. Social areas (including mostly cities, towns and villages) are mainly located in the coastal areas of the island (Fig. 2).

SAMPLING PROTOCOL

Survey localities were selected using a stratified sampling scheme based on two sources of information (land use and environmental variables) that were equally considered (detailed description in Aranda *et al.*, 2011). Thus, each survey locality represents a unique combination of five land uses (native forests, exotic forests, intensive pastures, semi-natural pastures and social areas) and four main environmental regions (R1, R2, R3 and R4) delimited by a cluster analysis, including climatic, topographical and geological variables (Fig. 2). Briefly, R1 is characterized by higher temperatures and low precipitation, whereas the highest annual rainfall and the lowest temperatures occur in R4. The regions R2 and R3 have relatively similar climatic environments, but the former is located in a lower and flatter altitudinal zone. We could not survey the sampling stratum corresponding to R1 × semi-natural pasture because all existing localities showed land use modifications. Accordingly, we surveyed a total of 19 localities trying to cover as much environmental range as possible with the amount of effort that we could invest in the field and in the laboratory (Table 1). This survey effort was decisive in deciding the number of localities to be surveyed and in previously determining the number of strata based on the land uses and environmental regions (see Aranda *et al.*, 2011).

In each selected locality, ground bryophytes were collected from five sampling zones separated by approximately 25 m along a 100 m × 10 m transect. If available, we sampled three replicates (vegetation plots) for each one of the considered substrates (rocks, soil and rotting wood) in each sampling zone. Hence, the number of vegetation plots per survey locality ranged from 15 to 33 (see Table 1). In each vegetation plot, we recorded all species and visually estimated their percentage cover (seven categories: presence < 1%, 1%, 1–5%, 5–20%, 20–50%, 50–75% and 75–100%). As the abundance of bryophytes differed greatly between land uses, we considered different plot sizes to obtain representative and comparable cover estimates in forests (30 cm × 30 cm), pastures (100 cm × 100 cm in soil and 30 cm × 30 cm in rocks) and social areas (5 cm × 5 cm). Most fieldwork was performed by the first and last authors over 34 days during the summer of 2008.

The specimens were identified at the species level using mainly Smith (2004) and Casas *et al.* (2006) for mosses and Paton (1999), Schumaker & Vána (2005) and Casas *et al.* (2009) for liverworts and hornworts. The species classification follows Goffinet, Buck & Shaw (2009) for mosses, Crandall-Stotler, Stotler & Long (2009) for liverworts and Renzaglia, Villarreal & Duff (2009) for hornworts. Doubtful taxa (66
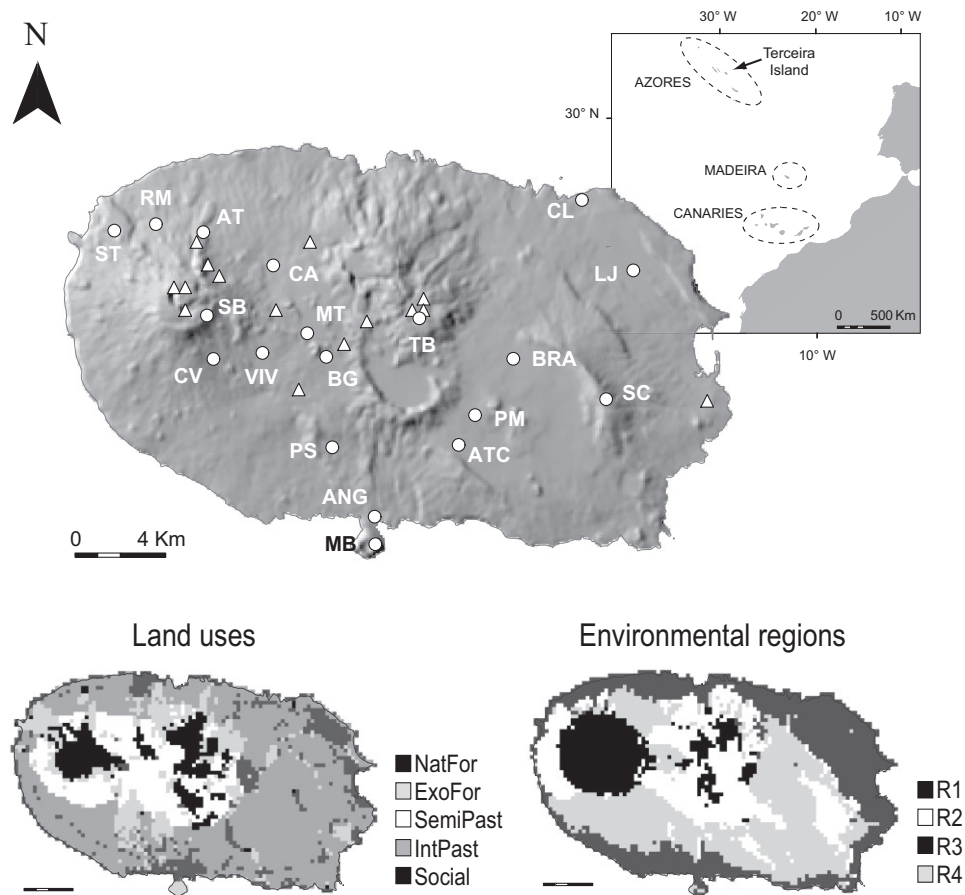
Figure 2. Geographical location of Terceira Island within the Macaronesian region, and topographic image of the island showing the 19 survey localities from the current field sampling (circles) and the 13 localities considered as well sampled according to prior knowledge (triangles). The distribution of land uses and environmental regions on the island is also shown. Codes of survey localities, land uses and environmental regions as in Table 1 (see Methods for further details).

specimens) were sent to specialists for taxonomic confirmation. The final database included 1944 records, which included collection information for each specimen found in the vegetation plots (available on request). Forty-one specimens from this dataset were identified at the genus level because of their damaged condition or lack of sexual characters. Voucher specimens have been deposited in the Herbarium of the University of Azores (AZU). For nomenclature, see Supporting Information Table S1.

DATA ANALYSES

Inventory completeness and sampling efficiency
We used species accumulation curves (SACs) to estimate the level of inventory completeness based on the number of species collected at each survey locality as a function of the performed sampling effort. The number of vegetation plots was used as a measure of sampling effort to build a smooth SAC (rarefaction

curve) after randomly resampling the plot order 1000 times (Gotelli & Colwell, 2011). From such rarefaction curves, the mean value of inventory completeness was calculated using the $\hat{C}$ estimator (Chao & Jost, 2012). $\hat{C}$ is related to the final slope of the SAC, and hence species inventories are considered to be more complete at higher $\hat{C}$ or lower slope values.

The sampling efficiency was examined by exploring the relationship between inventory completeness and sampling effort to then estimate the optimum number of vegetation plots needed to obtain comparable inventories. To do this, we assessed the statistical significance of the difference between the maximum inventory completeness reached at each survey locality and the equivalent value obtained with fewer vegetation plots. We calculated the overlap between the 84% confidence intervals for the mean completeness estimate as a statistical test with an α level of 0.05 (MacGregor-Fors & Payton, 2013). We then used an ANOVA test to calculate the main effects of the two

Table 1. Species richness (S) and number of effective species (exponential Shannon, e^H) found in each one of the 19 survey localities. e^H was calculated using the frequency of occurrence (i.e. number of vegetation plots in which each species was found). Each survey locality corresponded to unique combinations of the main land uses (LU: NatFor, native forests; ExoFor, exotic forests; SemiPast, semi-natural pastures; IntPast, intensive pastures; Social, social areas) and environmental regions (ER: R1–R4) existing in Terceira Island. The locality corresponding to R1 × semi-natural pasture could not be surveyed in the field (see text). Differences in species composition (Simpson dissimilarity index, β_{sim}) between the 500 m² Universal Transverse Mercator (UTM) well-surveyed cells (WSCs) selected from prior knowledge and current survey localities are also indicated. For each survey locality, the available substrates (soil, S; rock, R; rotting wood; RW), percentage of inventory completeness (IC) obtained with all sampling units (SU, i.e. number of vegetation plots) and the sampling effort that would have been needed to reach this same level of inventory completeness (Opt SU) are also shown. Inventory completeness was calculated with $\hat{C}$ using rarefaction curves (Chao & Jost, 2012), from which the mean ± 84% confidence intervals are provided (see text)

LU	ER	WSCs	Survey locality	Code	Substrates	SUs	S	e^H	β_{sim}	IC	Opt SUs
NatFor	R1	–	Caldeira das Lajes	CL	R, S	24	27	21.54	–	96.07 ± 2.50	16
NatFor	R2	4	Terra Brava	TB	R, RW, S	33	52	35.98	0.10	96.86 ± 1.12	25
NatFor	R3	6	Serra S. Bárbara	SB	R, S	27	60	44.75	0.14	95.72 ± 1.34	16
NatFor	R4	1	Matela	MT	R, S	20	42	26.66	0.41	92.65 ± 2.06	9
ExoFor	R1	–	Monte Brasil	MB	R, RW, S	27	15	10.59	–	96.83 ± 2.33	15
ExoFor	R2	–	Serreta	ST	R, RW	24	23	16.36	–	96.32 ± 2.07	15
ExoFor	R3	–	Altares	AT	R, RW	22	28	18.56	–	91.75 ± 2.86	10
ExoFor	R4	–	Posto Santo	PS	R	15	32	21.01	–	88.70 ± 3.44	6
SemiPast	R2	1	Pico da Bagacina	BG	R, S	22	29	18.38	0.33	85.72 ± 4.25	12
SemiPast	R3	–	Catarina Vieira	CV	R, S	18	17	13.11	–	86.91 ± 7.39	12
SemiPast	R4	1	Caminho dos Altares	CA	R, S	23	18	12.28	0.33	82.53 ± 8.25	8
IntPast	R1	–	Lajes	LJ	S	15	6	4.30	–	91.67 ± 7.43	5
IntPast	R2	–	Serra do Cume	SC	S	15	4	2.23	–	90.66 ± 5.22	3
IntPast	R3	–	Raminho	RM	S	15	6	4.27	–	97.41 ± 4.46	7
IntPast	R4	–	Pico Malhão	PM	S	15	3	2.52	–	100 ± 6.52	7
Social	R1	–	Angra	ANG	R, S	19	11	9.19	–	86.32 ± 9.67	9
Social	R2	–	Aterro-Canis	ATC	R	15	17	12.07	–	72.85 ± 9.86	4
Social	R3	–	Viveiro dos Florestais	VIV	R	15	18	14.64	–	88.82 ± 6.05	9
Social	R4	–	São Brás	BRA	R	15	14	10.32	–	82.76 ± 7.77	5

factors, 'land use' and 'environmental region', on the variation in the optimum sampling effort and inventory completeness among survey localities. In this way, we assessed whether survey efficiency differed between the levels of these two factors. We checked the normality of the residuals and the homoscedasticity assumptions. The statistical significance of all pairwise *post hoc* comparisons was assessed with a Tukey honestly significant difference (HSD) test adapted for unequal sample sizes between groups.

Previous and current knowledge of bryophyte species diversity

To examine the consistency between diversity estimates using recent field data and historical records coming from heterogeneous sources, we compared our sampling results with those obtained previously and made available in the literature and unpublished records (Aranda *et al*., 2010, 2011). We followed the most recent checklist of Azorean bryophytes (Gabriel

et al., 2010) with some updates (Aranda *et al*., 2013). In particular, we examined whether species assemblages remained similar between the former and present inventories using the 13 Universal Transverse Mercator (UTM) grid cells of 500 m × 500 m that had previously been identified as well surveyed and that may be comparable with our study [herein well-surveyed cells (WSCs), Aranda *et al*., 2011; see Fig. 1]. If available, we kept information on ground bryophyte data from previous collections. We used the Simpson index (β_{sim}) to estimate the assemblage resemblance based on presence–absence data, because this index is independent of variations in richness values (Lennon *et al*., 2001); β_{sim} ranges from zero to unity, higher values indicating greater dissimilarity.

Effects of the selected stratifying variables on bryophyte species diversity

To estimate whether the two stratifying variables (land use and environmental region) had a similar

effect on the variation in diversity among survey localities, we first identified the same level of inventory completeness at which comparisons could be feasible. We followed the methodology proposed by Chao & Jost (2012) for species richness (S), the exponential Shannon (e^H) and beta diversity (β_{sim}) by bootstrapping the community matrix 500 times and selecting the number of vegetation plots necessary to reach the minimum value of comparable completeness ($\hat{C}$). We hence evaluated the effect of land use on the variation in diversity among localities belonging to the same environmental region, and vice versa, by calculating the degree of overlap between the distributions of the bootstrapped values. Plot size did not affect the results here because comparisons of diversity were made regardless of sampling effort (see Chao & Jost, 2012). All computations were performed using STATISTICA (StatSoft Inc., 2013), the *vegan* package in R (Oksanen *et al.*, 2013) and iNEXT software (Hsieh, Ma & Chao, 2013).

RESULTS

INVENTORY COMPLETENESS AND SAMPLING EFFICIENCY

In total, we sampled 379 vegetation plots, from which we recorded 146 bryophyte species belonging to 50 families and 98 genera (Supporting Information Table S1). Survey localities could be considered well sampled because most exceeded 90% of inventory completeness and only one had < 80% of existing species collected (Table 1). However, similar inventory completeness could have been achieved with much less survey effort (Table 1).

Although no significant differences were found in sampling efficiency among environmental regions (not shown), differences among land uses were statistically significant in the optimum sampling effort ($F_{4,14} = 4.63$, $P = 0.014$) and inventory completeness ($F_{4,14} = 6.88$, $P = 0.003$). On average, it should be necessary to sample 17 vegetation plots in native forests, 12 in exotic forests, 11 in semi-natural pastures, six in intensive pastures and seven in social areas (Fig. 3A). Inventory completeness was statistically lower in social areas than in all the other land uses, except for semi-natural pastures (Fig. 3B).

PREVIOUS AND CURRENT KNOWLEDGE OF SPECIES DIVERSITY

In general, the species assemblages derived from the historical information and those obtained with the current field sampling remained relatively similar. All comparisons between the 13 WSCs and the five comparable survey localities gave $\beta_{sim} < 0.5$ (Table 1). In Serra S. Bárbara (SB) and Terra Brava (TB), 83% and
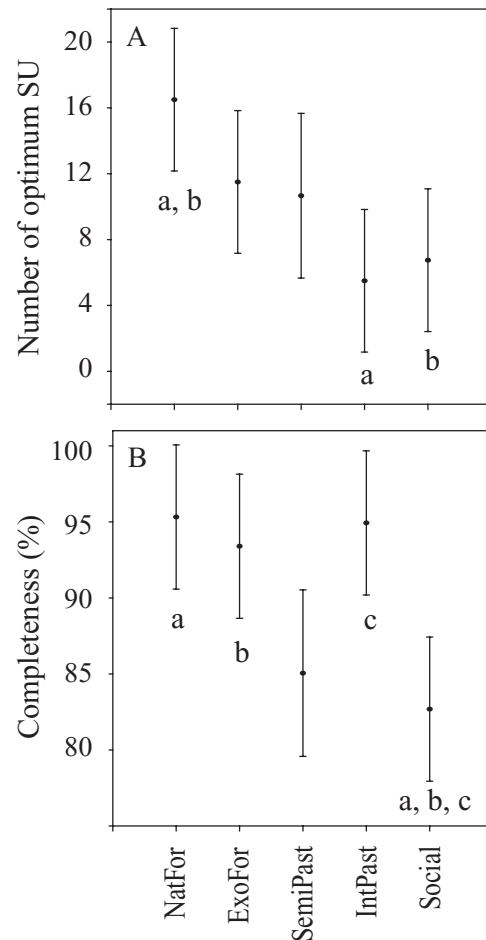

Figure 3. Variation (mean ± SE) in the amount of optimum sampling units (SU) (A) and inventory completeness (B) between land uses. Categories with the same letter indicate statistically significant differences ($P < 0.05$) according to a *post-hoc* Tukey's honestly significant difference (HSD) test adapted for unequal sample size between groups. Codes of land uses as in Table 1.

90%, respectively, of all the collected species had already been recorded in one or more of the WSCs within the stratum. For Matela (MT), Pico da Bagacina (BG) and Caminho dos Altares (CA), β_{sim} values oscillated from 0.33 to 0.41, as there was only one WSC in the whole stratum.

We recorded approximately 40% of all the species cited for Terceira Island and one-third of the species recognized for the whole archipelago. Three species were recorded for the first time in the Azores [*Rhynchostegiella litorea* (De Not.) Limpr., *Didymodon umbrosus* (Müll.Hal.) R.H.Zander and *Trichodon cylindricus* (Hedw.) Schimp.; see Ellis *et al.*, 2011, 2013], and five species were new citations for Terceira Island [*Ptychostomum torquescens* (Bruch & Schimp.) Ros & Mazimpaka, *Dicranella howei* Renauld &
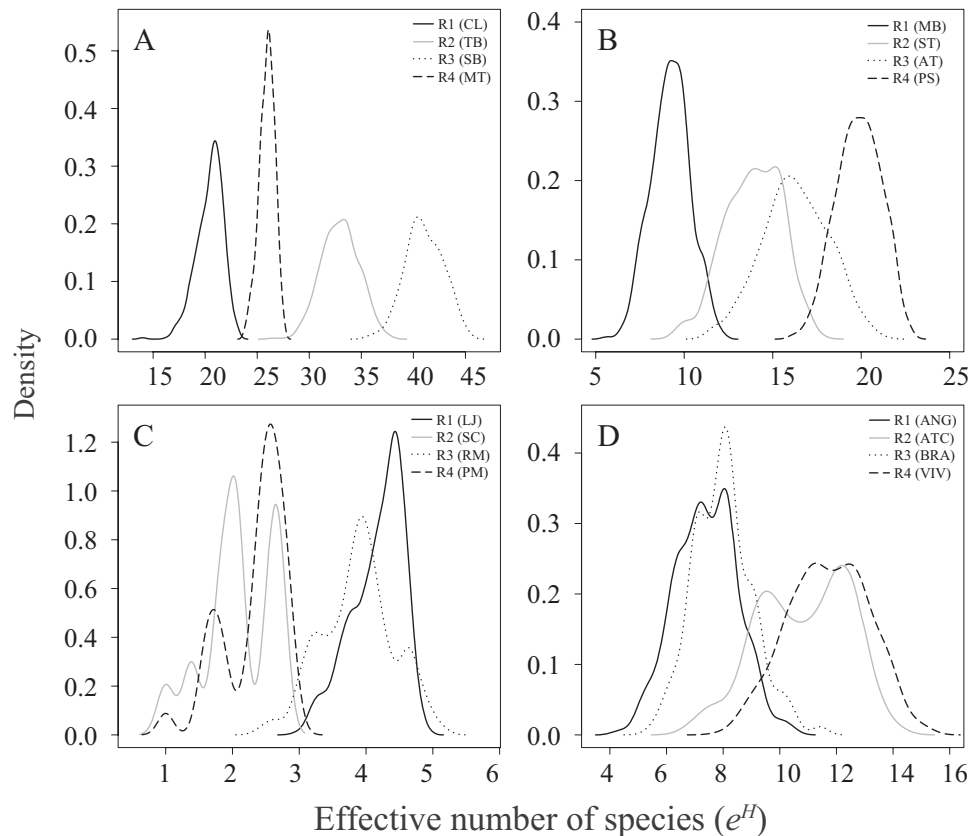
Figure 4. Comparisons of species diversity (exponential Shannon, e^H) between environmental regions in native forests (A), exotic forests (B), intensive pastures (C) and social areas (D). There was no significant relationships in the case of semi-natural pastures (see Supporting Information Table S2). e^H was calculated for each locality by bootstrapping the community matrix 500 times, choosing the number of samples (i.e. vegetation plots) necessary to reach the minimum value of comparable completeness in each case (see Table 1). The degree of overlap between the density distributions of such bootstrapped values was used to test significant differences in e^H (Supporting Information Table S2). The key shows the survey localities (in parentheses) corresponding to each environmental region (codes as in Table 1; see Methods for further details).

Cardot, *Southbya tophacea* (Spruce) Spruce, *Tortella tortuosa* (Hedw.) Limpr. and *Fissidens curvatus* Hornsch.]. Most of these new Azorean citations were collected in disturbed habitats that presented only one or no WSCs according to prior knowledge (Aranda *et al.*, 2011). For those species already known in the island, there was at least one new citation in 94% of cases; even for certain well-known common species, such as *Hypnum uncinulatum* Jur., *Andoa berthelotiana* (Mont.) Ochyra and *Campylopus flexuosus* (Hedw.) Brid., there could be up to five new localities. Furthermore, information about geographical location improved the precision for 12 species [e.g. *Brachythecium rivulare* Schimp., *Bryum dichotomum* Hedw., *Corsinia coriandrina* (Spreng.) Lindb., *Tortula solmsii* (Schimp.) Limpr.). Even in those cases with precise known locations, i.e. 500-m² UTM cells (Aranda *et al.*, 2010), there were no records from the last 30 years for

14 species, eight of them collected before 1985 [e.g. *Barbula unguiculata* Hedw., *Brachytheciastrum velutinum* (Hedw.) Ignatov & Huttunen, *Didymodon luridus* Hornsch.] and two others collected just once in 1937 [*Didymodon tophaceus* (Brid.) Lisa and *Tortula muralis* Hedw.).

EFFECTS OF THE SELECTED STRATIFYING VARIABLES ON SPECIES DIVERSITY

We focused our analyses on the number of effective species (e^H) because it was highly correlated with species richness (Spearman $r = 0.98$, $P < 0.001$; see Table 1). Our findings showed that most comparisons of e^H between environmental regions were statistically significant (Fig. 4; Supporting Information Table S2), except in the case of semi-natural pastures, where only the differences between R2 (Pico da
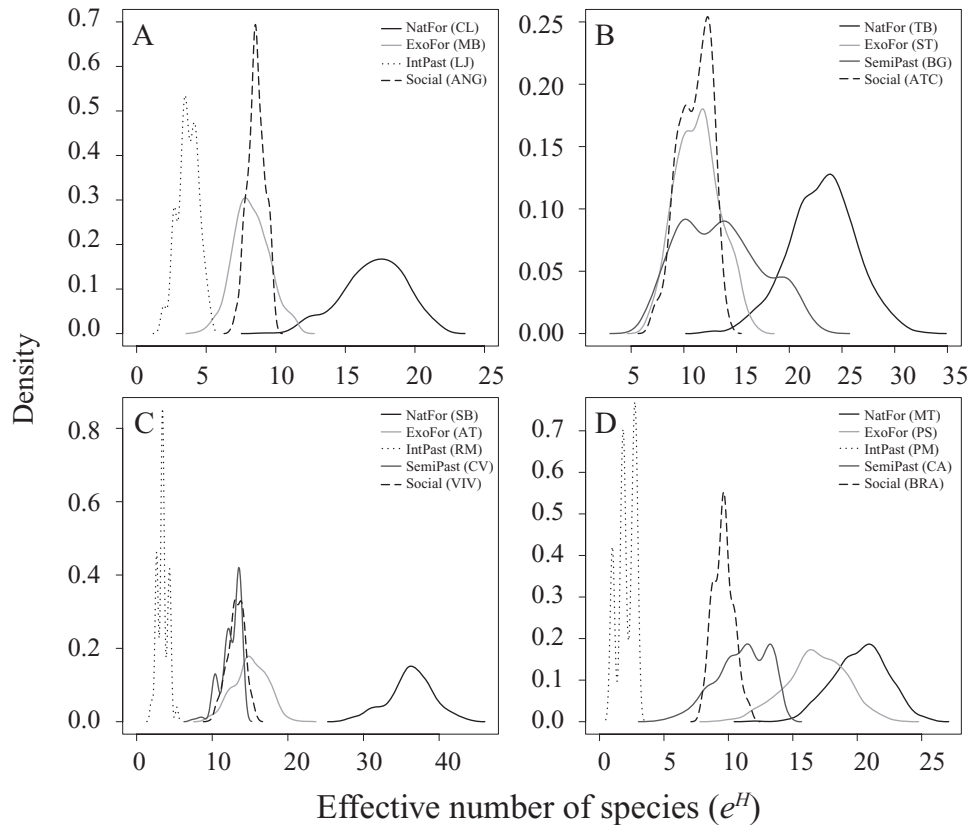
Figure 5. Comparisons of species diversity (exponential Shannon, e^H) between land uses in the four environmental regions: R1 (A), R2 (B), R3 (C) and R4 (D). See also Supporting Information Table S3. Further details as in Figure 4.

Bagacina) and R4 (Caminho dos Altares) remained marginally significant ($P = 0.06$). In forest, for instance, such differences may reach almost twice the diversity between distinct environmental regions. It should also be noted that the diversity values of R1 and R4 always differed significantly (Fig. 4). The comparisons among land use for each environmental region showed an obvious gradient from the richest native forests to the least diverse intensive pastures, with the other land uses in an intermediate position regardless of the environmental region (Fig. 5; Supporting Information Table S3). As expected, the magnitude of such differences was higher than those found in the analogous case among environmental regions (Fig. 4). However, the effective number of species in the forests of R4 (Fig. 5D) was not significantly different, regardless of their origin (either native or exotic). We also found that, among disturbed land use, semi-natural pastures, social areas and even exotic forests had similar numbers of effective species living on ground substrata.

Consistent results were obtained for species composition (Table 2; Supporting Information Tables S4, S5), which showed, on average, a higher dissimilarity

of local assemblages between land use belonging to the same environmental region (up to $\beta_{sim} = 0.84$) than vice versa (up to $\beta_{sim} = 0.71$). The value of community dissimilarity in the case of semi-natural pastures was much lower ($\beta_{sim} = 0.45$), in accordance with previous results on the number of effective species (Supporting Information Table S2).

DISCUSSION

This work completes a series which illustrates the difficulties and caution that need to be considered during the inventory process in any region or group of organisms (see Aranda *et al.*, 2010, 2011): (1) detection of taxonomic and geographical gaps; (2) planning of a sampling design to reduce the lack of knowledge previously detected; and (3) evaluation of the sampling results in relation to previous decisions for survey planning (Fig. 1). The latter is probably the most difficult step to be implemented because it is time consuming and usually requires interdisciplinary collaboration, especially when working with hyperdiverse groups at large geographical scales. It is clear from this that criticisms of the 'taxonomic

Table 2. Variation in species composition (Simpson dissimilarity, β_{sim}) between environmental regions in each land use, and vice versa. Multiple-site version of Simpson dissimilarity (Baselga, 2010) was calculated in each case by bootstrapping 500 times the minimum number of sampling units needed at equivalent levels of inventory completeness (see text for further details). Categories of land use and environmental regions as in Table 1 (see Methods for further details)

	Mean ± SD	Range (min–max)
Land use		
NatFor	0.67 ± 0.01	0.07 (0.64–0.70)
ExoFor	0.61 ± 0.02	0.10 (0.56–0.65)
SemiPast	N/A	N/A
IntPast	0.70 ± 0.03	0.17 (0.60–0.77)
Social	0.71 ± 0.03	0.16 (0.64–0.80)
Environmental region		
R1	0.84 ± 0.02	0.10 (0.81–0.91)
R2	0.83 ± 0.02	0.16 (0.74–0.90)
R3	0.79 ± 0.01	0.08 (0.75–0.82)
R4	0.73 ± 0.02	0.13 (0.66–0.78)

Because of the lack of survey locality corresponding to R1 × SemiPast, sample size limitations ($N = 3$) prevented us from calculating correlations in the case of semi-natural pastures (N/A, but see Supporting Information Table S4). Similarly, the number of pairwise comparisons in R1 was lower ($N = 6$) than in the other environmental regions.

impediment' should be correctly contextualized (Sluys, 2013). Hence, the iterative evaluation of the sampling results is a key aspect for the optimization of the survey design if the aim is to obtain reliable biodiversity patterns (e.g. Gaspar *et al.*, 2014). Using, as a case study, the bryophytes of a relatively small region with a great number of historical records, the evaluation of our survey planning suggests that: (1) we may improve our knowledge of regional diversity, but it may still be far from complete; (2) local values of inventory completeness may be high but, to be efficient, the survey effort should be unevenly distributed according to land use conditions; (3) the historical floristic information may be useful to validate and propose new field sampling protocols aimed at describing biodiversity patterns; and (4) the sources of information used in the stratified survey design should be weighted differentially according to the contribution in explaining diversity differences.

Without considering the species growing on trees, we have collected approximately 40% of the bryophyte diversity known in Terceira in < 1% of the island area sampled, and we have recorded the presence of eight species on the island for the first time. However, there are still obvious gaps of knowledge that must be covered. This is a clear example that species inventorying is, by nature, a complex process that must be continuously implemented, even when there is a large amount of historical information, as we show here for the bryophytes of Terceira (cf. Robertson, 2008; Boakes *et al.*, 2010; Beck *et al.*, 2013). The need for more taxonomic work is a frequent claim (Scheffers *et al.*, 2012). However, the key aspect of this study is that it quantifies and addresses how we should continue this work in the future. In particular, we call attention to the efficiency of the collecting protocol and to the performance of the variables chosen to select the survey localities according to prior knowledge of species ecology (Braun & Reynolds, 2012; Ferrer-Paris *et al.*, 2013). The underlying assumptions behind such decisions are limited by the working effort that we could invest, and should thus be examined to improve the sampling design and, ultimately, our understanding of biodiversity patterns (Hortal, Lobo & Jimenez-Valverde, 2007; Ballesteros-Mejia *et al.*, 2013).

We obtained high values of inventory completeness in most sampling localities, but we always exceeded the sampling effort needed. In fact, in many cases, we could have collected half of the vegetation plots to obtain the same species diversity. Moreover, both the optimum sampling effort and the inventory completeness varied as a function of the land use, but not the environmental region. Although these findings are context dependent, they suggest that collection protocols may not be sufficiently efficient, and that the variables chosen to plan the field survey in a region may differentially affect the sampling effort needed at a local scale (Coddington *et al.*, 2009; Watson, 2010). We should probably have distributed the sampling effort between, rather than within, localities, especially for disturbed places, such as social areas and intensive pastures, which showed lower values of inventory completeness. However, at the other extreme, we found that, in localities with lowest completeness values, the optimum sampling effort was also overcome (Table 1). These results point to the need for complementary protocols to capture the missing local rarities in these localities, such as ephemeral or spatially restricted species, for which expert knowledge from taxonomists is essential (Colwell & Coddington, 1994; Preston *et al.*, 2010). Hence, intensive local sampling also provides valid information that otherwise could not have been revealed.

The environmental diversity framework has shown contradictory results in capturing high species diversity (e.g. Araújo, Densham & Humphries, 2003; Medina *et al.*, 2013). In our case, it performed relatively well because there was a seeming effect of both stratifying variables, giving support to our prior hypotheses for survey planning (Aranda *et al.*, 2011).

However, the type of land use contributes more than the environmental regions to create differences in species diversity, suggesting that this issue should be further considered when sampling bryophytes (cf. Newmaster *et al*., 2005; Löhmus, Löhmus & Vellak, 2007; Corrales *et al*., 2010). Although the importance of land use for bryophyte diversity is well known by bryologists, to our knowledge, there are no previous studies in the literature showing the relative importance of this factor over environmental variables. Regardless from these results, we can say that differences in local diversity were statistically significant for some land uses and environmental regions, but more survey localities per stratum would be needed to support the effectiveness of the stratifying protocol itself (Hirzel & Guisan, 2002).

LESSONS FOR FUTURE PROSPECTS

This work did not aim to find the best sampling method of species diversity for a particular case study. It was intended as a first approach to orient the fieldwork when prior knowledge is lacking or insufficient, a situation that happens in many taxa and regions. Certainly, our results serve as a baseline to improve future sampling of bryophytes in Terceira (and other islands) and to study diversity–environment relationships (see Gabriel & Bates, 2005). However, we conclude by stressing three more general recommendations. (1) Whether or not previous information about species occurrences generates diversity values consistent with current inventories is a key question for the inclusion of such past data in biodiversity studies (Feeley & Silman, 2011; Lintz, Gray & McCune, 2013). (2) It is desirable to evaluate the effectiveness of the invested sampling effort in obtaining reliable inventories and representative data of biodiversity variation. This sampling effort should ideally be established according to the spatial and environmental variation prevailing in the study area, and limited by the experience of species collection and identification and the time and resources available to carry out this work. (3) The main factors influencing species distribution should be chosen before planning the fieldwork according to expert taxonomists, but their relevance should also be assessed after carrying out the survey in order to estimate reliable biodiversity patterns.

ACKNOWLEDGEMENTS

REFERENCES

Aranda SC, Gabriel R, Borges PAV, de Azevedo EB, Lobo JM. 2011. Designing a survey protocol to overcome the Wallacean shortfall: a working guide using bryophyte distribution data on Terceira Island (Azores). *The Bryologist* **114:** 611–624.

Aranda SC, Gabriel R, Borges PAV, Lobo JM. 2010. Assessing the completeness of bryophytes inventories: an oceanic island as a case study (Terceira, Azorean archipelago). *Biodiversity and Conservation* **19:** 2469–2484.

Aranda SC, Gabriel R, Borges PAV, Santos AMC, Hortal J, Baselga A, Lobo JM. 2013. How do different dispersal modes shape the species–area relationship? Evidence for between-group coherence in the Macaronesian flora. *Global Ecology and Biogeography* **22:** 483–493.

Araújo MB, Densham P, Humphries C. 2003. Predicting species diversity with ED: the quest for evidence. *Ecography* **26:** 380–383.

Ballesteros-Mejia L, Kitching IJ, Jetz W, Nagel P, Beck J. 2013. Mapping the biodiversity of tropical insects: species richness and inventory completeness of African sphingid moths. *Global Ecology and Biogeography* **22:** 586–595.

Baselga A. 2010. Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography* **19:** 134–143.

Basset Y, Cizek L, Cuénoud P, Didham RK, Guilhaumon F, Missa O, Novotny V, Ødegaard F, Roslin T, Schmidl J, Tishechkin AK, Winchester NN, Roubik DW, Aberlenc H-P, Bail J, Barrios H, Bridle JR, Castano-Meneses G, Corbara B, Curletti G, da Rocha WD, de Bakker D, Delabie JHC, Dejean A, Fagan LL, Floren A, Kitching RL, Medianero E, Miller SE, de Oliveira EG, Orivel J, Pollet M, Rapp M, Ribeiro SP, Roisin Y, Schmidt JB, Sorensen L, Leponce M. 2012. Arthropod diversity in a tropical forest. *Science* **338:** 1481–1484.

Bebber DP, Marriott FHC, Gaston KJ, Harris SA, Scotland RW. 2007. Predicting unknown species numbers using discovery curves. *Proceedings of the Royal Society B: Biological Sciences* **274:** 1651–1658.

Beck J, Ballesteros-Mejia L, Nagel P, Kitching IJ. 2013. Online solutions and the 'Wallacean shortfall': what does GBIF contribute to our knowledge of species' ranges? *Diversity and Distributions* **19:** 1043–1050.

Beck J, Boeller M, Erhardt A, Schwanghart W. 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics* **19:** 10–15.

Boakes EH, McGowan PJK, Fuller RA, Ding CQ, Clark NE, O'Connor K, Mace GM. 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biology* **8:** e1000385.

Braun DC, Reynolds JD. 2012. Cost-effective variable selection in habitat surveys. *Methods in Ecology and Evolution* **3:** 388–396.

Brown JH, Lomolino MV. 1998. *Biogeography*. Sunderland, MA: Sinauer Associates.

Cardoso P. 2009. Standardization and optimization of arthropod inventories – the case of Iberian spiders. *Biodiversity and Conservation* **18:** 3949–3962.

Casas C, Brugués M, Cros RM, Sérgio C. 2006. *Handbook of mosses of the Iberian Peninsula and the Balearic Islands: illustrated keys to genera and species*. Barcelona: Institut d'Estudis Catalans.

Casas C, Brugués M, Cros RM, Sérgio C, Infante M. 2009. *Handbook of liverworts and hornworts of the Iberian Peninsula and the Balearic Islands*. Barcelona: Institut d'Estudis Catalans.

Chao A, Jost L. 2012. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* **93:** 2533–2547.

Chapman AD. 2009. *Numbers of living species in Australia and the world*. Canberra: Australian Biodiversity Information Services.

Chytry M, Tichy L, Hennekens SM, Schaminee JHJ. 2014. Assessing vegetation change using vegetation-plot databases: a risky business. *Applied Vegetation Science* **17:** 32–41.

Coddington JA, Agnarsson I, Miller JA, Kuntner M, Hormiga G. 2009. Undersampling bias: the null hypothesis for singleton species in tropical arthropod surveys. *Journal of Animal Ecology* **78:** 573–584.

Colwell RK, Coddington JA. 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* **345:** 101–118.

Corrales A, Duque A, Uribe J, Londono V. 2010. Abundance and diversity patterns of terrestrial bryophyte species in secondary and planted montane forests in the northern portion of the Central Cordillera of Colombia. *The Bryologist* **113:** 8–21.

Crandall-Stotler B, Stotler RE, Long DG. [2008] 2009. Morphology and classification of the Marchantiophyta. In: Goffinet B, Shaw AJ, eds. *Bryophyte biology, 2nd edn*. Cambridge: Cambridge University Press, 1–54.

Croft MV, Chow-Fraser P. 2009. Non-random sampling and its role in habitat conservation: a comparison of three wetland macrophyte sampling protocols. *Biodiversity and Conservation* **18:** 2283–2306.

Dengler J. 2009. A flexible multi-scale approach for standardised recording of plant species richness patterns. *Ecological Indicators* **9:** 1169–1178.

Ellis LT, Asthana AK, Gupta R, Nath V, Sahu V, Bednarek-Ochyra H, Ochyra R, Cykowska B, Aranda SC, Fischer E, Gabriel R, Gorski P, Gremmen N, Hespanhol H, Kurbatova LE, Smith RIL, Long DG, Bell D, Mogro F, Sergio C, Garcia CA, Stow S, Martins A, Smith VR, Vana J, Vanderpoorten A. 2013. New national and regional bryophyte records, 34. *Journal of Bryology* **35:** 62–70.

Ellis LT, Bednarek-Ochyra H, Ochyra R, Calvo Aranda S, Colotti MT, Schiavone MM, Dulin MV, Erzberger P, Ezer T, Kara R, Gabriel R, Hedenas L, Holyoak DT, Odor P, Papp B, Sabovljevic M, Seppelt RD, Smith VR, Sotiaux A, Szurdoki E, Vanderpoorten A, van Rooy J, Zarnowiec J. 2011. New national and regional bryophyte records, 26. *Journal of Bryology* **33:** 66–73.

Faith DP, Walker PA. 1996. Environmental diversity: on the best-possible use of surrogate data for assessing the relative biodiversity of sets of areas. *Biodiversity and Conservation* **5:** 399–415.

Feeley KJ, Silman MR. 2011. Keep collecting: accurate species distribution modelling requires more collections than previously thought. *Diversity and Distributions* **17:** 1132–1140.

Ferrer-Paris JR, Rodriguez JP, Good TC, Sanchez-Mercado AY, Rodriguez-Clark KM, Rodriguez GA, Solis A. 2013. Systematic, large-scale national biodiversity surveys: NeoMaps as a model for tropical regions. *Diversity and Distributions* **19:** 215–231.

Funk VA, Richardson KS, Ferrier S. 2005. Survey-gap analysis in expeditionary research: where do we go from here? *Biological Journal of the Linnean Society* **85:** 549–567.

Gabriel R, Bates JW. 2005. Bryophyte community composition and habitat specificity in the natural forests of Terceira, Azores. *Plant Ecology* **177:** 125–144.

Gabriel R, Sjögren E, Schumacker R, Sérgio C, Aranda SC, Claro D, Homem N, Martins B. 2010. List of bryophytes (Anthocerotophyta, Marchantiophyta, Bryophyta). In: Borges PAV, Costa A, Cunha R, Gabriel R, Gonçalves V, Martins AF, Melo I, Parente M, Raposeiro P, Rodrigues P, Santos RS, Silva L, Vieira P, Vieira V, eds. *A list of the terrestrial and marine biota from the Azores*. Cascais: Princípia, 99–115.

Gaspar C, Borges PAV, Gaston KJ. 2008. Diversity and distribution of arthropods in native forests of the Azores archipelago. *Arquipélago. Life and Marine Sciences* **25:** 1–30.

Gaspar C, Cardoso P, Borges PAV, Gaston KJ. 2014. Efficiency of sampling methods and effort to assess arthropod diversity in Azorean native forests. *Arquipélago. Life and Marine Sciences* **31:** 21–36.

Goffinet B, Buck WR, Shaw AJ. 2009. Morphology and classification of the Bryoophyta. In: Goffinet B, Shaw AJ, eds. *Bryophyte biology, 2nd edn*. Cambridge: Cambridge University Press, 55–138.

Gotelli NJ, Colwell RK. 2011. Estimating species richness. In: Magurran AE, McGill BJ, eds. *Frontiers in measuring biodiversity*. New York: Oxford University Press, 39–54.

Hirzel A, Guisan A. 2002. Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling* **157:** 331–341.

Hortal J, Lobo JM, Jimenez-Valverde A. 2007. Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands. *Conservation Biology* **21:** 853–863.

Hsieh TC, Ma KH, Chao A. 2013. iNEXT online: interpolation and extrapolation (Version 1.0) 1.0 ed. Available at: http://chao.stat.nthu.edu.tw/inext/

Lennon JJ, Koleff P, Greenwood JJD, Gaston KJ. 2001. The geographical structure of British bird distributions: diversity, spatial turnover and scale. *Journal of Animal Ecology* **70:** 966–979.

Lintz HE, Gray AN, McCune B. 2013. Effect of inventory method on niche models: random versus systematic error. *Ecological Informatics* **18:** 20–34.

Löhmus A, Löhmus P, Vellak K. 2007. Substratum diversity explains landscape-scale co-variation in the species-richness of bryophytes and lichens. *Biological Conservation* **135:** 405–414.

MacGregor-Fors I, Payton ME. 2013. Contrasting diversity values: statistical inferences based on overlapping confidence intervals. *PLoS ONE* **8:** e56794.

Medina NG, Lara F, Mazimpaka V, Hortal J. 2013. Designing bryophyte surveys for an optimal coverage of diversity gradients. *Biodiversity and Conservation* **22:** 3121–3139.

Michalcova D, Lvoncik S, Chytry M, Hajek O. 2011. Bias in vegetation databases? A comparison of stratified-random and preferential sampling. *Journal of Vegetation Science* **22:** 281–291.

Mutke J, Geffert JL. 2010. Keep on working: the uneven documentation of regional moss floras. *Tropical Bryology* **31:** 7–13.

Newmaster SG, Belland RJ, Arsenault A, Vitt DH, Stephens TR. 2005. The ones we left behind: comparing plot sampling and floristic habitat sampling for estimating bryophyte diversity. *Diversity and Distributions* **11:** 57–72.

Oksanen J, Blanchet FG, Kindt R, Legendre L, Minchin R, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2013. vegan: community ecology package. R package version 2.0-7. Available at: http://CRAN.R-project.org/package=vegan

Paton AJ, Brummitt N, Govaerts R, Harman K, Hinchcliffe S, Allkin B, Lughadha EN. 2008. Towards Target 1 of the Global Strategy for Plant Conservation: a working list of all known plant species – progress and prospects. *Taxon* **57:** 602–611.

Paton JA. 1999. *The liverwort flora of the British Isles.* Leiden: Brill Academic Publishers.

Preston CD, Hill MO, Porley RD, Bosanquet SDS. 2010. Survey of the bryophytes of arable land in Britain and Ireland 1: a classification of arable field assemblages. *Journal of Bryology* **32:** 61–79.

Renzaglia KS, Villarreal JC, Duff RJ. 2009. New insights into morphology, anatomy, and systematics of hornworts. In: Goffinet B, Shaw AJ, eds. *Bryophyte biology, 2nd edn.* Cambridge: Cambridge University Press, 139–171.

Risely K, Renwick AR, Dadam D, Eaton MA, Johnston A, Baillie SR, Musgrove AJ, Noble DG. 2011. *The breeding bird survey 2010. BTO Research Report 597.* Thetford: British Trust for Ornithology.

Robertson DR. 2008. Global biogeographical data bases on marine fishes: caveat emptor. *Diversity and Distributions* **14:** 891–892.

Sánchez-Fernández D, Lobo JM, Abellán P, Millán A. 2011. How to identify future sampling areas when information is biased and scarce: an example using predictive models for species richness of Iberian water beetles. *Journal for Nature Conservation* **19:** 54–59.

Sánchez-Fernández D, Lobo JM, Abellán P, Ribera I, Millán A. 2008. Bias in freshwater biodiversity sampling: the case of Iberian water beetles. *Diversity and Distributions* **14:** 754–762.

Sastre P, Lobo JM. 2009. Taxonomist survey biases and the unveiling of biodiversity patterns. *Biological Conservation* **142:** 462–467.

Scheffers BR, Joppa LN, Pimm SL, Laurance WF. 2012. What we know and don't know about Earth's missing biodiversity (vol 27, pg 501, 2012). *Trends in Ecology and Evolution* **27:** 712–713.

Schumaker R, Vána J. 2005. *Identification keys to the liverworts and hornworts of Europe and Macaronesia: (distribution and status).* Poznań: Sorus.

Sluys R. 2013. The unappreciated, fundamentally analytical nature of taxonomy and the implications for the inventory of biodiversity. *Biodiversity and Conservation* **22:** 1095–1105.

Smith AJE. 2004. *The moss flora of Britain and Ireland.* Cambridge: Cambridge University Press.

Sousa-Baena MS, Garcia LC, Peterson AT. 2014. Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. *Diversity and Distributions* **20:** 369–381.

StatSoft Inc. 2013. STATISTICA (data analysis software system), version 8.0. Available at: http://www.statsoft.com

Watson DM. 2010. Optimizing inventories of diverse sites: insights from Barro Colorado Island birds. *Methods in Ecology and Evolution* **1:** 280–291.

Yoccoz NG. 2012. The future of environmental DNA in ecology. *Molecular Ecology* **21:** 2031–2038.

Zhang Z-Q. 2011. Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness. *Zootaxa* **3148:** 1–237.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Distribution of all species collected in the 19 sampling localities.

Table S2. Differences in species diversity (S and e^H) between environmental regions (ER) for each land use (LU).

Table S3. Differences in species diversity (S and e^H) between land uses (LU) for each environmental region (ER).

Table S4. Differences in species composition (Simpson dissimilarity, β_{sim}) between environmental regions (ER) in each land use (LU).

Table S5. Differences in species composition (Simpson dissimilarity, β_{sim}) between land uses (LU) in each environmental region (ER).