RESEARCH ARTICLE

Diversity and Distributions WILEY

The value of targeted biological surveys: An assessment of Australia's Bush Blitz programme

Chris Ware[1] | Kristen J. Williams[2] | Simon Ferrier[2] | Daniel P. Faith[3] | Glenn Manion[4] | Brian Hawkins[5] | Thomas D. Harwood[2] | Art Langston[2] | Jacqui Meyers[2] | Justin Perry[6] | Jo Harding[5] | Sue Fyfe[5]

[1]Environment, Commonwealth Scientific and Industrial Research Organisatiom (CSIRO), Hobart, Tasmania, Australia

[2]Environment, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra, Australian Capital Territory, Australia

[3]The Australian Museum Sydney, Sydney, New South Wales, Australia

[4]Cryzan Pty Ltd, Armidale, New South Wales, Australia

[5]Bush Blitz Program, Australian Biological Resources Study, Department of Climate Change, Energy, the Environment and Water, Australian Government, Canberra, Australian Capital Territory, Australia

[6]Environment, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Townsville, Queensland, Australia

Correspondence
Kristen J. Williams, Environment, Commonwealth Scientific and Industrial Research Organisation (CSIRO), GPO Box 1700, Canberra, ACT 2601, Australia.
Email: kristen.williams@csiro.au

Present address
Brian Hawkins, NSW Department of Planning and Environment, Coffs Harbour, New South Wales, Australia

Thomas D. Harwood, Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, UK

Abstract

Aim: Biodiversity assessment and decisions rely on knowledge of the spatial distribution of species, yet most global biodiversity is inadequately represented by occurrence records. Efforts to improve our knowledge of biodiversity distribution include targeted taxon survey programmes aimed at generating records of new, or previously unrecorded, species. Here, we evaluate nearly 8 years of biodiversity record collection by Bush Blitz, Australia's largest species discovery programme, to test how efficiently knowledge was added through the programme.

Location: Continental Australia.

Methods: Because we expect locations that are environmentally distinct in comparison with those already surveyed to harbour novel records of species, we assess the extent to which Bush Blitz surveys complement continental environmental diversity (ED). We then assess how effectively this improvement in sampling of ED translates into the accumulation of records of new, or previously unrecorded, species. Our assessment is based on Bush Blitz data for six taxa (amphibians, spiders, land snails, moths and butterflies, reptiles and vascular plants), benchmarked against data accumulated over the same period by the Atlas of Living Australia (ALA)—Australia's largest aggregation of biodiversity records—as a comparison of effectiveness to the Bush Blitz programme.

Results: Environments surveyed through the Bush Blitz programme are highly complementary to environments from which 'background' observations were made over the same period and aggregated in the ALA. Bush Blitz surveys result in large numbers of records of new, or previously unrecorded, species. Across most biological groups considered, additions were made highly efficiently with respect to survey effort, relative to background survey effort represented in the ALA.

Main Conclusions: Our results demonstrate the ability of the Bush Blitz programme to contribute valuable data to conservation assessment and planning and the important

Chris Ware and Kristen J. Williams are considered as co-lead authors.

role of surrogate-based assessments of ED complementarity in planning new targeted biological surveys.

1 | INTRODUCTION

Continued collection of biological data through national field-survey programmes is essential to improving knowledge of biodiversity and underpins our ability to adequately manage natural areas into the future (Brito, 2010; Reddy, 2010). Globally, the biodiversity of most regions is only partially documented, and this is particularly the case for less well-studied taxonomic groups (Oliver et al., 2021). For example, in Australia, it is estimated that only around one-third of all species have been documented, with the undocumented portion heavily biased towards groups such as invertebrates (Chapman, 2009; Taxonomy Decadal Plan Working Group, 2018). New surveys, coupled with the entry of existing survey data into publicly accessible national repositories such as the Atlas of Living Australia (ALA) (Belbin et al., 2021; Belbin & Williams, 2016) and the Global Biodiversity Information Facility (GBIF Secretariat, 2019), should improve our capacity to manage biodiversity and make informed decisions. Yet, quantitative assessments of how well new surveys have improved knowledge of overall biodiversity and, by extension, capacity to manage and conserve biodiversity, are rarely undertaken (Balmford & Gaston, 1999; Carvalho et al., 2010; Grantham et al., 2008; Lõhmus et al., 2018).

Existing assessments of the value of investing in biodiversity data collection present a mixed picture. Balmford and Gaston (1999) found new biodiversity surveys to be of high value due to the efficiency with which survey data could be used to aid in the spatial prioritisation of new protected areas. Meanwhile, Grantham et al. (2008) concluded that, when other dimensions of conservation are considered (e.g. limited money to spend, challenges in filling data gaps efficiently and adequacy of existing data), ongoing survey efforts yield limited value. Other studies have found either support for additional data collection (Chowdhury et al., 2023; Hanson et al., 2023; Ware et al., 2018) or mixed (Field & Elphick, 2019; Rodewald et al., 2019) or limited support (Maxwell et al., 2015; Mentges et al., 2021). Ascertaining the general value of collecting additional biodiversity data from published tests, therefore, is not straightforward. However, the complex picture presented exists largely due to differences in methods and data used among studies, how and for what purpose data were collected, the conservation objectives tested, taxonomic scope, and whether socioeconomic factors were considered. In most cases, it is difficult to separate the effect of how complete biological datasets are from other confounding factors (Kujala et al., 2018). Increasingly, methods are being developed to test the effectiveness and improve the utility of additional data collection

(Callaghan et al., 2019; Hanson et al., 2023; Mesaglio et al., 2023). The value of additional data to conservation planning will certainly decline at some point simply because the incremental difference made reduces as the dataset becomes more complete (Possingham et al., 2007). In contrast, the value of additional data for more poorly documented groups should be substantial (Troudet et al., 2017), but assessments for these groups are lacking.

Methods to assess the value of additional biodiversity data for achieving conservation goals have mostly used tests of reserve selection based on biological datasets of different size and complexity (e.g. Beier & de Albuquerque, 2016; Coetzee et al., 2014; Dunn et al., 2016; Gaston & Rodrigues, 2003; Rodewald et al., 2019). Given the value of additional data should be greatest where existing data are most incomplete, a different but complementary means of assessing the value of collecting additional data lies in tests of existing data completeness. 'Environmental diversity' (ED) (Faith, 1994, 2003; Faith & Walker, 1996a) is a biodiversity surrogates framework for linking species data and environmental gradients that is well suited to such tests. Several empirical tests of the ED framework demonstrate that it performs well when used to assess biodiversity representativeness (Beier & de Albuquerque, 2015, 2016; Engelbrecht et al., 2016). The framework assumes species exhibit unimodal responses to environmental gradients (i.e. high abundances of any given species are concentrated in some portion of environmental space) (Faith, 2003), implying that sites which are distant in terms of their environmental characteristics will support different species.

Measures, therefore, of how well the ED of a region is represented by locations where species surveys have been conducted, should, as a result, translate into an estimate of how adequately the species diversity of a region has been sampled (Ferrier, 2002). Faith and Walker (1996a) demonstrated that, under this model, the number of species sampled will be maximised if a set of survey locations is distributed such that, on average, the distance from any point in environmental space to the nearest survey site is minimised. This application to biodiversity measurement is a special form of the p-median problem (Church, 2002; ReVelle & Swain, 1970). It follows that summing the distance from each and every location in a region to the most similar existing survey site provides a measure of diversity 'forgone' (Faith & Walker, 1996a) and can be interpreted when calculated for a given region and set of survey sites, as an estimate of survey completeness (see Appendix S1.1 for a description of the p-median problem).

The original multivariate biodiversity surrogate formulation used by Faith and Walker (1996a) to demonstrate the concept of

ED has since been extended by the introduction of generalised dissimilarity modelling (Faith et al., 2004; Ferrier et al., 2002). Generalised dissimilarity modelling (GDM) (Ferrier et al., 2007) is a statistical technique for analysing and predicting spatial patterns in species assemblage data that flexibly accommodates nonlinearity in relationships between patterns of species occurrence and environmental gradients. As a surrogate for conservation decision-making, biologically scaled environmental gradients derived from GDMs have been used to assess, for example gaps in biological survey data (Ashcroft et al., 2010; Bell et al., 2014), protected-area representativeness (Ferrier et al., 2004; Williams et al., 2016) and the proportion of species expected to be retained within a region as a function of different configurations of natural habitat (Allnutt et al., 2008; Di Marco et al., 2019; Ferrier et al., 2020; Mokany et al., 2020).

Here, we present analyses of the relative efficiency of targeted taxon surveys using the Bush Blitz programme as a test case. Bush Blitz (www.bushblitz.org.au) is Australia's largest species discovery programme. It is a multiyear, multimillion-dollar partnership between the Australian Government, BHP and Earthwatch Australia (https://earthwatch.org.au/), developed with the aim of recording new information, including new species to science, new occurrence records, and range extensions (Preece et al., 2015). One of the overarching objectives of the programme is to fill gaps in the sampling of ED. Our assessment uses data from 8 years of these expeditionary surveys (2009–2017) to evaluate how the programme: (1) improves coverage of ED represented in biological record aggregations; and (2) efficiently generates records of new, or previously unrecorded, species relative to other survey efforts based on national aggregations available through the Atlas of Living Australia.

We consider six taxonomic groups in our assessment (amphibians, spiders, land snails, moths and butterflies, reptiles and vascular plants) and determine the value of additional data collection separately for each. In undertaking these analyses, we also provide additional evidence for the general utility of the ED framework as a surrogate of species diversity.

2 | MATERIALS AND METHODS

2.1 | Analytical overview

The main components of our analysis are presented in Figure 1. The backbone of our analytical framework is comprised of models of biologically scaled environmental space. These are derived by fitting generalised dissimilarity models (GDM: Ferrier et al., 2007) to available observation records for each taxonomic group considered, and a suite of covariates characterising biologically relevant environmental gradients. To assess both the degree to which Bush Blitz contributed to sampling ED and how this equated to new records of species, we compared the performance of Bush Blitz to a separate dataset referred to as 'background' data. We considered records available through the Atlas of Living Australia (ALA) as the background data

for two reasons. First, the ALA freely provides access to Australia's largest aggregation of records of biological observations regularly ingested from custodian natural history institutions and individual collectors. Second, the ALA's vision is to deliver trusted biodiversity data services for Australia supporting world-class research and decision-making (Belbin et al., 2021). Implicit in our analyses is the assumption that national conservation decisions are mostly made based on ALA data (or information sets derived from ALA data) and that the ALA is complete in the sense that all available data is accessible via that platform. Our analyses thus assess the efficiency with which additional information was collected by the Bush Blitz programme compared with the ALA over the same period.

Improvement in ED coverage was assessed by evaluating how well each new survey location complemented the existing set of surveyed locations in representing ED for a given taxon. These calculations were performed using a formulation of the p-median algorithm, and models of biologically scaled environments to represent ED. Species discovery (either new records to science, or records of species in new locations) was assessed by tallying the accumulation of new records as they were collected by date, or in order of their ED complementarity using the p-median algorithm. The latter approach iteratively identified which of the set of surveyed sites would best fill gaps in ED. When standardised by the total number of records collected on each survey date or at each site as selected by the p-median algorithm (described below), the relative rate of accumulation indicates the performance of the Bush Blitz programme in species discovery relative to background observations accessed via the ALA (excluding any ingested records attributable to the Bush Blitz programme). Both analyses were performed relative to the 'baseline' given as the background ALA dataset prior to the commencement of Bush Blitz in 2009.

2.2 | The Bush Blitz programme

The Bush Blitz programme was developed to fill gaps in knowledge of Australia's biodiversity (Preece et al., 2015). The programme undertakes a series of intense, targeted expeditions, where taxonomic experts representing diverse biological groups combine to carry out surveys of priority sites. The multidisciplinary team pool resources (e.g. a pitfall trap for vertebrates will also collect other target taxa such as invertebrates) and shares knowledge to increase data and specimen collection and to generate a better understanding of plant–animal relationships (e.g. Cheng & Cassis, 2019). Target taxa are usually selected to maximise species discovery for groups that have large numbers of undescribed species (e.g. spiders), together with better described groups (e.g. vascular plants) of conservation priority. However, each expedition focusses on a range of different taxa that can vary depending on taxon expert availability. The programme started in 2009 and had conducted 36 expeditions by 2018 (Figure 2) across continental Australia (grouped by timing or general location in Appendix Table S2.1). The programme has contributed upwards of 40,000
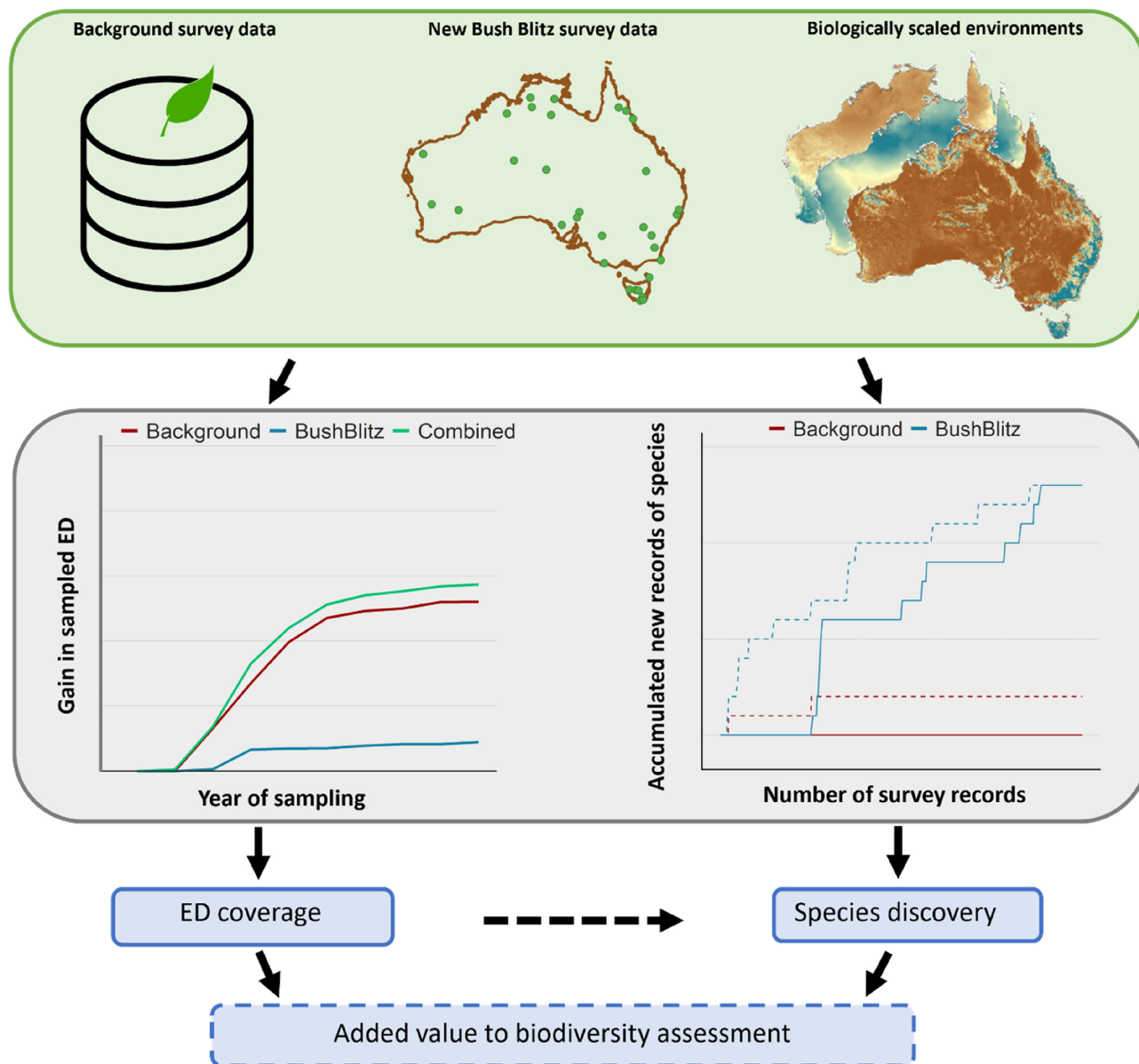
FIGURE 1 Overview of the analysis framework used to assess the efficiency with which Bush Blitz surveys sampled ED and contributed records of new species, or records of species at previously unrecorded locations, compared with background survey data drawn from the Atlas of Living Australia (ALA). Biologically scaled environments derived from generalised dissimilarity models of each assessed biological group underpinned the complementarity assessment of sampled ED (described in Ware et al. (2018) and summarised in Appendix Table S2.4). The relative performance of the Bush Blitz dataset, background dataset and their combination, in improving ED representation was calculated using a new formulation of the p-median algorithm (see main text for details). The ED measure is expressed in units of compositional dissimilarity, whereby increases in dissimilarity indicate additional survey locations have reduced the average dissimilarity of any location given a set of existing surveyed sites (depicted in the central-left plot of 'gains in sampled ED' vs. 'year of sampling'). Species discovery is assessed using two measures depicted in the central-right plot by solid and dashed lines. The order of accumulation of new records of species is tallied either by date (solid line) or by ED complementarity using the p-median algorithm (dashed line). Accumulated new records of species are standardised by the total number of records collected ordered by their observation date or by ED complementarity according to the p-median algorithm and thus represent a measure of efficiency.

new biological occurrence records representing 39 biological groups, including records of more than 1800 putative new species (http://bushblitz.org.au/statistics/).

The choice of Bush Blitz survey locations depends on several criteria. Foremost among these is the prospect of species discovery in one or more of the target taxa, weighed against more

practical considerations such as land manager engagement opportunities and infrastructure to support the expedition (Preece et al., 2015). Bush Blitz expeditions initially focussed on Australian National Reserve System properties (DAWE, 2021; National Reserve System Task Group, 2009). Since 2014, site selection expanded to include land managed under some form of conservation

FIGURE 2 Location of Bush Blitz expeditions (orange dots) conducted between 2009 and 2017. Details of the surveys are provided in Appendix Table S2.1. Numbers correspond to the ID column in Table S2.1. Basemap: Bing Maps Aerial (Microsoft product screen copy reprinted with permission from Microsoft Corporation).

agreement which includes national parks, Indigenous Protected Areas and private land (Preece et al., 2015). ED was used as a proxy for expected species discovery and calculated for target taxa to identify optimal survey locations. Resulting maps of unsampled ED were intersected with knowledge from taxonomic experts and incorporated in the expedition site selection process. This process was updated annually with data location from completed expeditions.

Specimens collected by taxon experts are held by natural history institutions in the jurisdiction where they were collected or, by agreement, held in a partner institution. Following taxonomic determination, specimen records are digitised and mobilised into the ALA. This process, from observation to public repository, can take many years depending on the scientific staff and resources of the

custodian institution and so Bush Blitz offers financial support to facilitate expedited curation of specimens, publication of new species and digitisation.

2.2.1 | Biological survey data

Background biological survey data for each of the six taxa were obtained from the ALA (extracted 16th January 2018) using custom R scripts adapted from the ALA4R package (Newman et al., 2019; Raymond et al., 2014; described in Appendix Section S1.2.1). The ALA data were split into two sets: records before the first Bush Blitz survey in September 2009 (ABRS, 2009) and records after the 36th expedition in May 2017 (ABRS, 2017a, 2017b). The actual

start and end months varied slightly depending on the taxon. Records were only considered in our analyses if those taxa observations were made on, or prior to, the 36th Bush Blitz expedition (i.e. amphibians: 17 May 2017; Araneae: 8 May 2017; land snails: 16 March 2017; Lepidoptera, reptiles; vascular plants 18 May 2017).

All biological records sourced from the ALA were further subject to a set of filters to remove potentially spurious, or inaccurate records, or records of non-native species. Data quality assertions provided by the ALA (Belbin et al., 2020; Miles, 2011) aided decisions about which records to remove (see Appendix Section S1.2.2, and Tables S1.1 and S1.2).

Bush Blitz survey data for the 36 expeditions were sourced from the programme. Because some of these records had already been deposited in the ALA (excluding new or cryptic species yet to be described or identified and made publicly accessible), we identified and removed Bush Blitz records from the ALA data download, which is to be used to represent background survey effort method described in Appendix Section S1.2.3.

The ALA aggregates biological records collected in diverse ways, including observations and specimen data (organisms, images and recordings). Observations and specimens come from many sources, including museum and natural history collections, universities, indigenous ecological knowledge holders, science agencies, individuals, community and conservation groups, government and industry (ALA, 2023). Institutes collect and publish biological records for a range of reasons beyond species discovery. We used all records available in the ALA and did not attempt to filter according to how the data were collected or by who. High-level breakdowns of the data sources and record types of the ALA data that we used are provided for each taxon in Appendix Figures S3.1a and S3.1b.

2.2.2 | Surrogate environmental diversity patterns

GDMs of compositional turnover for four taxonomic groups (vascular plants, reptiles, amphibians and land snails), as previously described in Ware et al. (2018), were used as surrogate measures of ED. Each site was associated with the value of environmental predictor variables representing climatic, topographic, water balance and substrate gradients, as listed in appendix 6 of Reside et al. (2013). Initial models were fitted using all variables, and variable selection followed the approach outlined by Williams et al. (2012) to remove redundant variables and reduce model complexity (i.e. avoid overfitting). The indirect effect of geographical distance between site pairs was tested as an additional variable once a reduced model had been derived and was retained in the final model if its unique contribution was above the specified minimum. See Appendix Table S2.2 for details of predictors used in each taxon model. Because there were insufficient records available to fit models to either Araneae or Lepidoptera records, we used the vascular plant GDM as a surrogate

as shown by Ware et al. (2018) to perform reasonably well for these groups.

A key output of a GDM is the transformation of environmental predictors into values of biological importance placed on a common scale. The resulting values are in units of ecological distance (E), such that the difference in value between any two points (i,j) for a given variable (x) represents the contribution to ecological distance between the two points for that variable. By summing the absolute differences across all transformed variables ($|x_i - x_j|$, i.e. the Manhattan distance), we obtain the change in E (ΔE) between the two points. Applying a negative exponential back-transformation of the logistic link function used in the GDM, and then subtracting this from one, the ecological distances can be expressed as Sørenson compositional dissimilarity. A value of one represents the modelled expectation that all species are shared between sites, and values approaching zero indicate no species are in common. When continuous environmental grids are in this way biologically scaled, back-transformed and mapped, the resulting patterns can be interpreted as spatial variation in species composition as a function of the predictors used in the model. Because GDM-transformed grids are scaled according to their importance in representing patterns of biological turnover for the taxonomic group in question, they provide a valuable means of assessing survey coverage consistent with the ED framework. Therefore, we refer to these GDM-transformed environmental grids as 'biologically scaled'.

2.3 | Analysis of improvement in environmental diversity coverage

This analysis was applied to the six taxa, for each year of survey between September 2009 and May 2017 (except for reptiles, first surveyed November 2009). Four of the taxa were matched with ED patterns modelled using data for that taxon, and two others (Araneae and Lepidoptera) were matched with vascular plants (as outlined above).

The following analyses were undertaken for each taxon:

- Baseline ALA: Prior to the first survey conducted by the Bush Blitz programme, how much of Australia's ED had already been sampled?
- Bush Blitz: How much additional ED did the Bush Blitz survey programme sample above baseline each successive year (i.e. 2009, 2010, …, 2017)?
- Background ALA: How much additional ED did other observations (i.e. background records) sample above baseline during the same period as the Bush Blitz survey programme each successive year?
- Effective: Overall, how much more knowledge about Australia's ED did the Bush Blitz programme contribute, above that contributed by other observations each year?

2.3.1 | Adapting ED complementarity method for assessing representation of biodiversity by surveys

One of the primary applications of the ED framework is its use in prioritising locations for further species surveys (e.g. Funk et al., 2005). Doing so requires two elements (Faith & Walker, 1996b): first, a model of ED for all locations in a given study region; and, second, an optimisation procedure to maximise the span of ED covered by a new set of survey sites. Faith and Walker (1996b) argued that the *p*-median optimisation model, developed in the scientific field of facilities location (operations research), provided the most appropriate method to optimally locate new survey sites. In the context of prioritising sites for surveys, the *p*-median problem consists of locating new survey sites, such that the sum of distances from each of the unsurveyed sites to its nearest survey site is minimised. How well each site complements the existing surveyed sites in ED is referred to as ED complementarity (Faith & Walker, 1996b) and is given by the site's contribution to the *p*-median value, such that a site with high ED complementarity will lower the *p*-median value more relative to other potential sites (for further explanation, including our novel solution for representing continuous ED, see Appendix S1.1). Typically, this form of survey gap analysis is conducted using gridded representations of regions or countries and multidimensional environmental space.

To assess incremental improvements in continuous ED coverage, we used the *p*-median approach to calculate annual *p*-median values for the Bush Blitz, ALA background and combined datasets. The approach we implemented was performed using GDM-transformed grids and is outlined in Appendix Section S1.2. Individual *p*-median calculations for each grid cell of continental Australia were summed to derive a 'global *p*-median' for each spatial layer of the analysis design. The global *p*-median statistic reduces to 0 when the ED of the study region is fully represented by surveyed sites. Once transformed into units of compositional dissimilarity (as per Section 2.2.2), the global ED statistic measures the average proportion of species shared between any site and the set of surveyed sites. A value of one represents no shared species, and zero represents the same complement of species.

2.3.2 | Analysis of Bush Blitz efficiency in sampling environmental diversity

To evaluate the relative efficiency of Bush Blitz in sampling ED compared with ALA background records, we accumulated annual gains in ED complementarity using the global *p*-median statistic to show the rate in two ways: (1) plotted in chronological order by year of observation, and (2) the plots of chronological order standardised by cumulative number of additional sites (250 m grid cells) with observation records—as a measure of survey effort. We expect Bush Blitz records to accumulate faster, especially when standardised by survey effort. To measure achievement over the study period, we calculated the area under the curve (AUC). We also derived their combined rate to show how Bush Blitz records complemented ALA background (i.e. generally filled gaps in ED).

2.4 | Analysis of Bush Blitz efficiency in species discovery

We measured the efficiency of Bush Blitz expeditions in terms of generating records of new, or previously unrecorded, species relative to ALA background data using the same biological survey data subsets and questions as for improvements in ED coverage (Section 2.3): ALA baseline observations prior to the first Bush Blitz expedition, Bush Blitz expeditions and ALA background observations during the same period as the Bush Blitz surveys. ALA baseline data established which species were observed before the Bush Blitz programme started in September 2009. We used rates of new, or previously unrecorded, species accumulation as the measure of performance and accumulated surveyed sites and their species in two ways: (1) in the order given by their observation date; (2) by selecting the site which iteratively best complemented the ED represented by the existing set of surveyed sites according to the *p*-median algorithm. We expect the rate of new, or previously unrecorded, species accumulation by Bush Blitz expeditions to always be higher than background observations, and for the optimal accumulation rate based on *p*-median ordering of complementarity to always be higher than chronological ordering. Locations with higher *p*-median values are more likely to yield different species because their environments are relatively distinct from those already sampled.

For method (2), we iteratively selected locations within a survey set in order of their contributions to filling gaps in ED space. Because each optimally selected location will often mean nearby locations do not contribute additional information (i.e. characterised by the same environments, and thus do not reduce the *p*-median further), we devised a new way to correctly associate those locations with the nearest optimally selected location in ED space and accordingly accumulate their species (see Appendix Section S1.3 and Tables S1.1 and S1.2 for details).

As each surveyed location is added, the number of new, or previously unrecorded, species were accumulated on the *y*-axis, while the number of all species records made (of either new, or previously unrecorded, species, or records of previously described species) are accumulated as a proxy for survey effort on the *x*-axis (i.e. termed 'survey event'). When plotted, the resulting species accumulation curves measure both success and efficiency in species discovery.

2.5 | Relationship between sampling of environmental diversity and new records of species across taxa

We expected that the number of records of new, or previously unrecorded, species contributed by new surveys for each taxon should be related to how much new ED was sampled between 2009 and 2017.

To explore this, we used simple linear models to evaluate whether decreases in average GDM-based ED covaried with numbers of new records of species. Numerous factors other than how well ED was sampled should affect this relationship, including how comprehensively a taxon was surveyed at baseline, and the overall species richness of the taxon. The purpose of our simple model was to provide supporting evidence for the general utility of the ED framework as a surrogate of species diversity applied to this analysis of Bush Blitz efficiency. To partially account for the factors listed above, we used the number of new, or previously unrecorded, species as a proportion of those records available at baseline as this somewhat accounts for previous survey effort and species richness. Linear models of the proportion of new species observed as a function of ED sampled were fit using the R statistical programming language (R Core Team, 2020).

3 | RESULTS

3.1 | Summary of biological survey data

For all taxa considered in our assessment, Bush Blitz surveys generated more records of new, or previously unrecorded, species than ALA background records over the study period since Bush Blitz commenced

(Table 1). On average, a record of a new, or previously unrecorded species was generated at every 1.7 sites (grid cells) or for every 11.9 observations of species made on Bush Blitz expeditions. For background survey data, it took on average 98.7 sites or 1399 observations for a record of a new, or previously unrecorded species. More new, or previously unrecorded, species records of Lepidoptera were discovered than for other taxa, with just over two new records discovered at every Bush Blitz analysis site on average. Moths, for example, are among one of the most species-rich insect groups in Australia and only about half of the estimated 22,000 species have been scientifically named (Austin et al., 2004; Zborowski & Edwards, 2007). High numbers of new records were also discovered of Araneae (also a species-rich group) and land snails through Bush Blitz expeditions, while the fewest records of new, or previously unrecorded species, were of amphibians (either through Bush Blitz expeditions or ALA background records; Table 1).

3.2 | Improvement in environmental diversity coverage and sampling efficiency

Improvements in ED coverage consistently led to reductions in average species compositional dissimilarity to surveyed locations above baseline for all taxa (Figures 3 and 4). The continent-wide gain in

TABLE 1 Summary of biological record data for the six taxonomic groups considered in our assessment.

Group	Data subset	*n* sites	*n* records	*n* species
Amphibians	Baseline	53,787	210,747	205
	Background	10,413	46,098	7
	Bush Blitz	272	957	13
Araneae	Baseline	3506	12,551	1164
	Background	948	2369	113
	Bush Blitz	452	1962	486
Gastropods (land snails)	Baseline	3507	12,550	1167
	Background	948	2370	113
	Bush Blitz	557	2291	416
Lepidoptera	Baseline	3548	24,826	1401
	Background	4471	27,691	872
	Bush Blitz	601	10,726	1213
Reptiles	Baseline	75,918	272,756	766
	Background	14,050	54,623	52
	Bush Blitz	663	2420	55
Vascular plants	Baseline	642,359	7,429,214	20,653
	Background	93,782	1,632,400	105
	Bush Blitz	2140	14,722	580
Total	Background	124,612	1,765,551	1262
	Bush Blitz	4685	33,078	2763

Note: The summary is broken down by Bush Blitz and background (ALA) records and the existing baseline (ALA prior to the start of Bush Blitz in 2009). '*n* sites' is the unique number of sites (9-arcsecond grid cells) with at least one species observation record; '*n* records' is the total number of observations records; and '*n* species' is the total number of species (richness). For Bush Blitz and ALA background data subsets, '*n* sites' and '*n* species' numbers are new (i.e. unique) relative to ALA baseline.
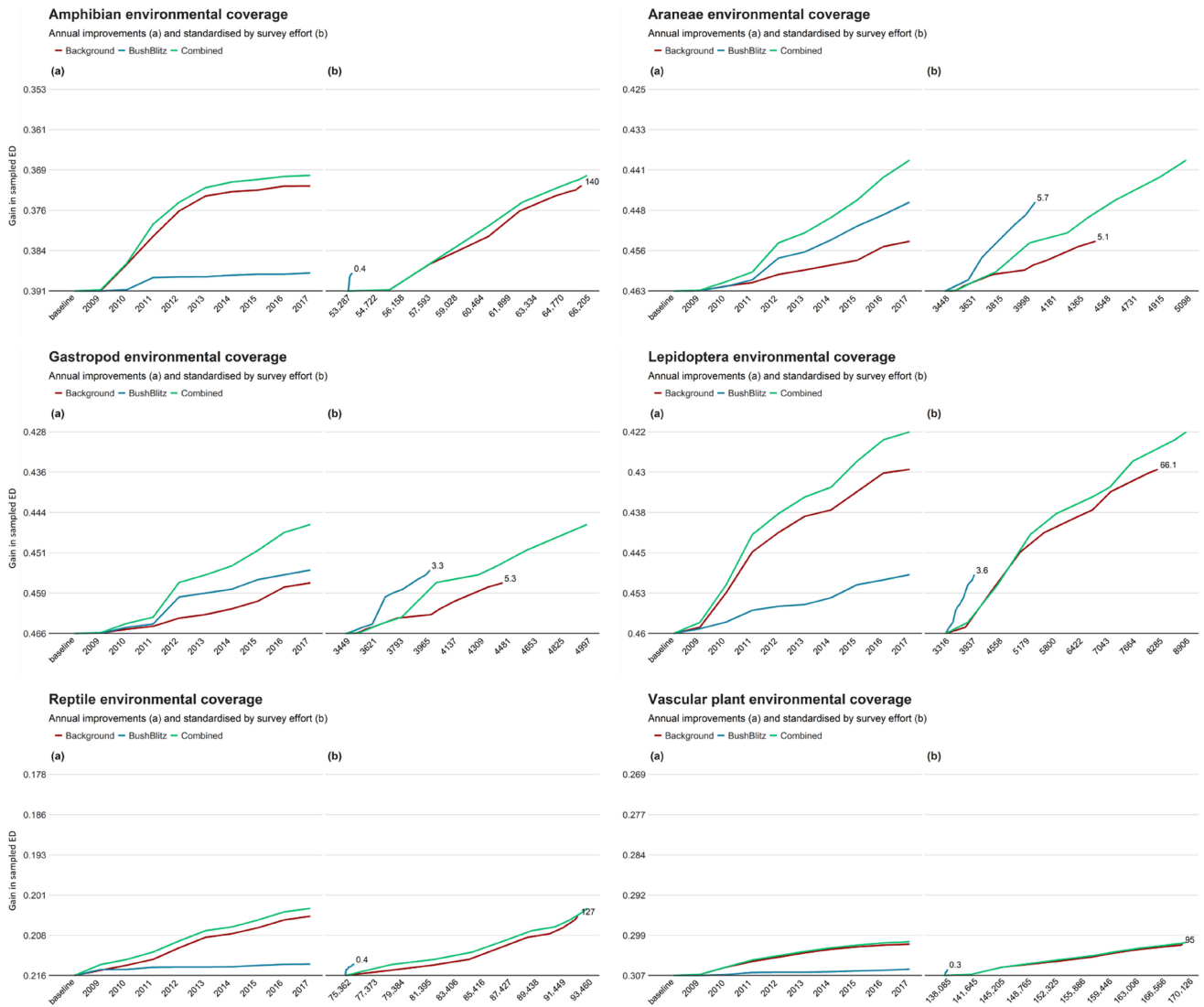
FIGURE 3 Annual gains in environmental diversity (ED) sampling above baseline for the six taxon groups between 2009 and 2017 attributed to either Bush Blitz (blue), ALA background (red) or their combination (green); presented in two ways: (a) cumulative annual gain in sampled ED and (b) annual gain standardised by survey efficiency—the cumulative number of sites surveyed each year (baseline is the first number of surveyed grids on the x-axis). The y-axis is stretched to the maximum range above the respective baseline across all six biological groups for comparability. Units of sampled ED (y-axis) are expressed as the average proportion of species not shared between any site and the set of surveyed sites based on the 'global p-median' statistic. The two numbers on panel (b) displayed at the peaks of the Bush Blitz (blue) and ALA background (red) lines, are the area under the curve (AUC) statistics as a measure of the total relative survey effort for that improvement in ED.

ED coverage attributed to Bush Blitz, over that contributed by ALA background records over the same period, varies for each of the six taxa (Figure 3 and Table S2.3). Considering just the chronology of surveys (Figure 3a, panels), ALA background records sampled more gaps in ED than Bush Blitz expeditions, except for spiders and land snails. However, when survey effort is considered, Bush Blitz expeditions were always more efficient (Figure 3b, panels). For amphibians, for example, there were few Bush Blitz sites (272) compared with ALA background (53,787), yet Bush Blitz expeditions rapidly sampled new ED, whereas for the equivalent survey effort ALA records revealed only a negligible gain in ED. When Bush Blitz and ALA background records are combined (green lines in Figure 3), the

result is additive, indicating survey effort is generally complementary—that is rarely overlaps sites with similar ED.

The complementary nature of Bush Blitz and ALA background records (2009 to 2017) in contributing to gains in ED coverage above baseline is also apparent spatially in Figure 4. Each new species observation record (or cluster of records within a 250m grid cell) potentially contributes to gains in ED according to how distinct the biologically scaled environment is relative to all other unsurveyed locations above baseline. The spatial extent of any gain in ED resulting from a single surveyed location is a function of how restricted or broad the environment is in which the survey location exists. If the surrounding environment is heterogeneous, numerous
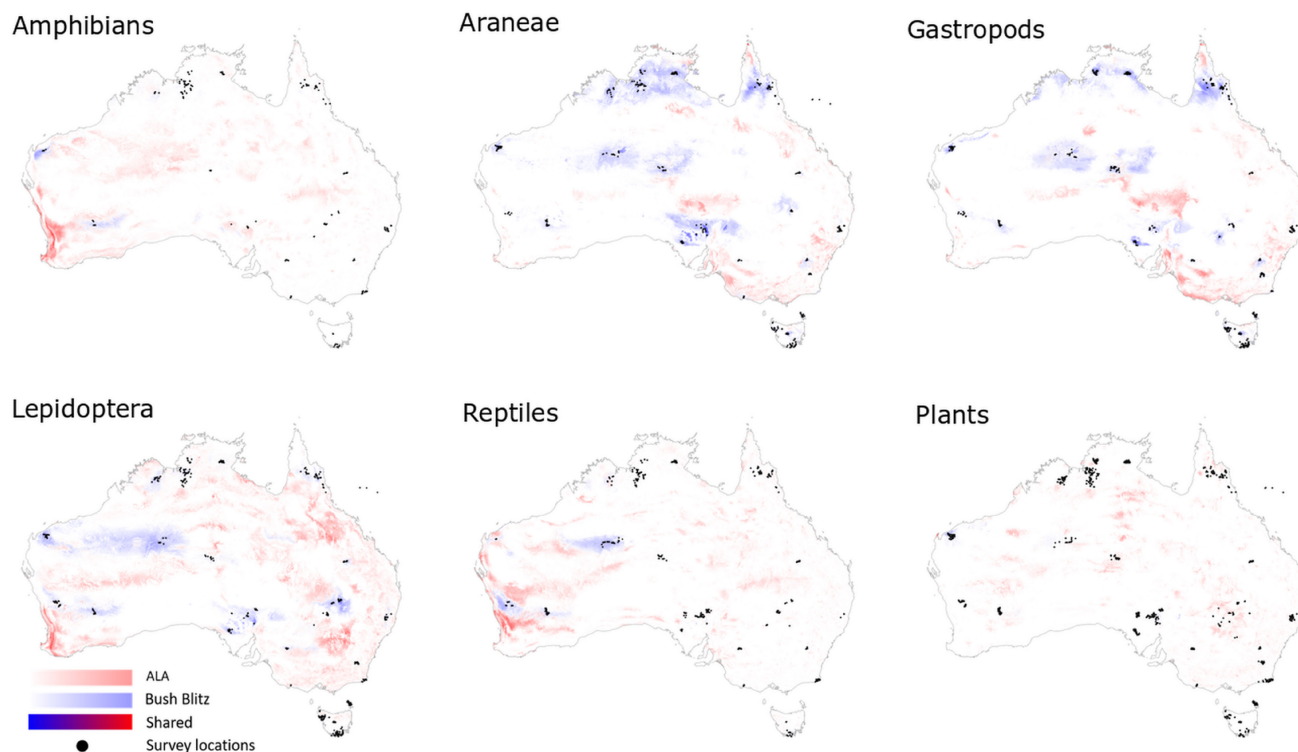
FIGURE 4 Regional gains in environmental diversity (ED) coverage above baseline for the six taxa uniquely attributed to either Bush Blitz (blue shades) or ALA background (red shades) records, or shared by both Bush Blitz and ALA records (equal contributions are represented by violet shades), between 2009 and 2017 for continental Australia. Lighter hues indicate less gain relative to baseline and darker hues greater gain in ED complementarity. Where both ALA and Bush Blitz records overlap in contributing to regional gains in ED, their relative influence is indicated by how reddish (ALA) or bluish (Bush Blitz) the resulting violet colour is. Gain in ED complementarity is calculated as the shortest Manhattan distance between an individual location (250 m pixel) and any existing surveyed location in ED space. Black dots are Bush Blitz survey locations for the respective group.

survey locations may be required to adequately sample regional ED, whereas in a homogenous environment a single survey location may suffice to adequately sample regional ED. Because the three invertebrate groups (Araneae, Lepidoptera and land snails) were among the least well-surveyed taxa across continental Australia at baseline, the additional surveys between 2009 and 2017 had the greater spatial gains in ED (depicted by the greater spread of blue shades for Bush Blitz and red for ALA background records in Figure 4). In contrast, the gains in ED for relatively well-known taxa such as vascular plants are more localised, where remaining gaps are in more spatially restricted environments.

The ALA baseline survey effort reveals knowledge gaps across the western and arid interior of Australia and remote localities of northern Australia for most taxonomic groups (Appendix Figures S3.2–S3.7). Some of these gaps were filled by specific Bush Blitz expeditions, which individually accounted for a reduction in ED of between 0.12% for vascular plants and 1.68% for Araneae, in units of average species dissimilarity (mean across the six taxa is 0.78%; Appendix Table S2.3). Background observations on the other hand accounted for a reduction of between 0.59% for vascular plants and 3.10% for Lepidoptera in average species dissimilarity (Appendix Table S2.3). Improvements in ED coverage contributed by ALA background or Bush Blitz varied by taxa: Bush Blitz

improvements were greatest for Araneae and land snails, while ALA background improvements were greatest for Lepidoptera and amphibians (Figure 4, Appendix Table S2.3).

3.3 | Efficiency with which new records of species were accumulated

For all taxa, Bush Blitz expeditions generated more records of new, or previously unrecorded, species and at a much faster rate (i.e. more efficiently) than ALA background records (Figure 5). This was evident irrespective of whether new records of species were tallied chronologically, or in the order they sampled gaps in ED (using the p-median algorithm). However, records of new, or previously unrecorded, species were more often accumulated faster when the order of sites from which additional records were tallied was determined by the p-median algorithm irrespective of the Bush Blitz or ALA background source. The order of Bush Blitz sampling for three taxa (Araneae, Lepidoptera and vascular plants) by date or by ED complementarity appeared to make less difference to the rate of accumulation of records of new, or previously unrecorded, species (less separation in the curves). These were the more species rich of the six groups analysed. The accumulation rate of records of new, or
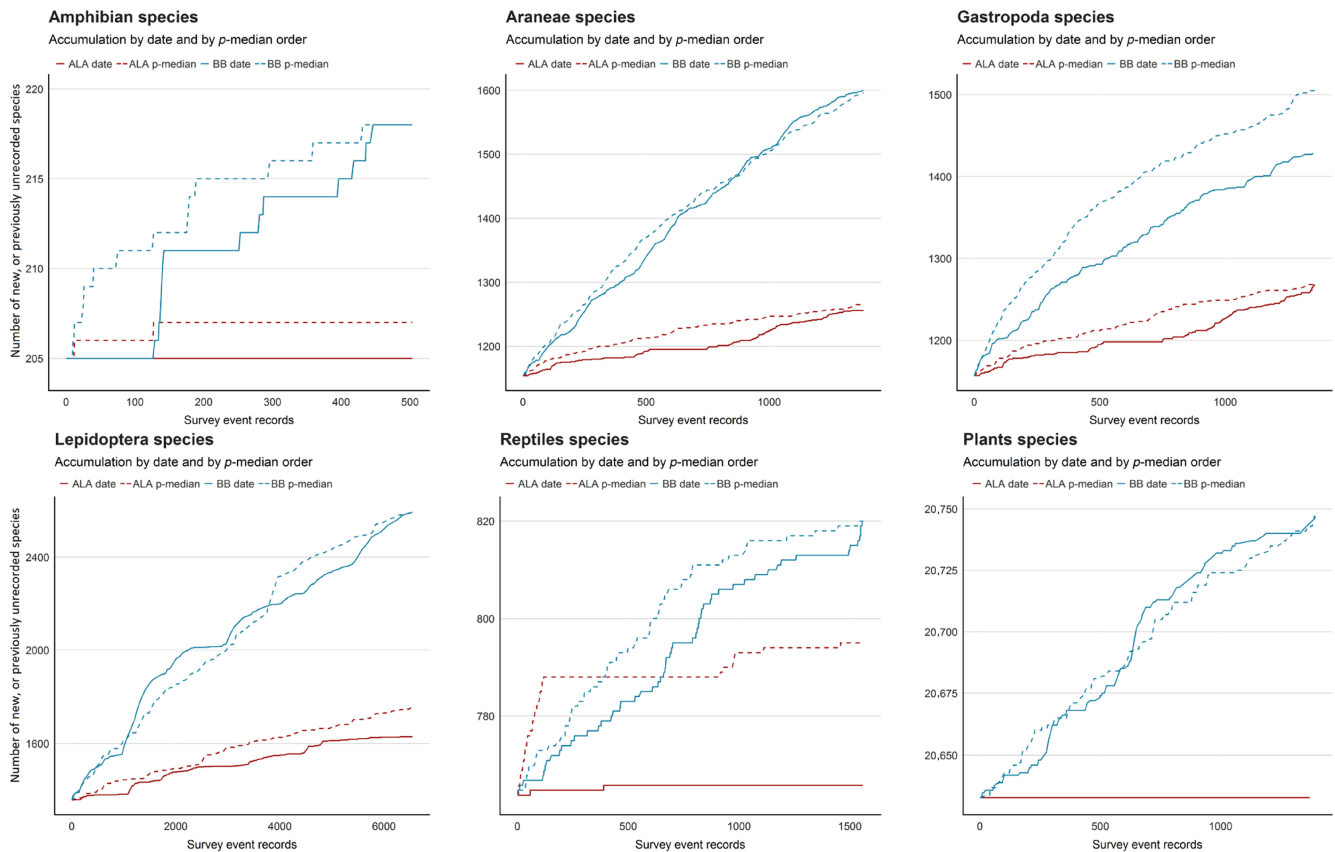
FIGURE 5 Gains in species discovery (species accumulated as new event records were added) for the six biological groups between 2009 and 2017, presented in two ways: in chronological order by observation date (solid lines) and by ED complementarity order according to the *p*-median algorithm (dotted lines). Accumulation is scaled by survey event records—a measure of the number of species records associated with a survey event (unique site and date), as a proxy for survey effort. The *x*-axis range is determined by the minimum number of survey event records from either the Bush Blitz or ALA background records. Bush Blitz records are shown in blue, ALA background in red. ALA background records include systematic surveys as well as ad hoc observations.

previously unrecorded, species is sensitive to how comprehensively each group was surveyed at baseline and at each survey location subsequently, and the Bush Blitz targeted survey methodology did not necessarily aim to achieve comprehensive sampling of all potential species at all sites.

3.4 | Relationship between environmental diversity and new records of species across taxa

We found only weak relationships between the incremental sampling of ED and resulting numbers of records of new, or previously unrecorded, species (ALA background: $p = <.05$, $R^2 = .72$. Bush Blitz: $p = .11$, $R^2 = .50$, Figure 6). In the case of Bush Blitz data, results for Lepidoptera strongly influenced the weak relationship, whereas amphibian data weakened the relationship for ALA background data. Amphibians are represented by fewer species than the other taxa analysed, and Lepidoptera are among the most species rich. When these taxa were omitted as outliers from the respective model, the relationships strengthened markedly (ALA background: $p = <.001$, $R^2 = .98$ excluding amphibians; Bush Blitz:

$p = <.01$, $R^2 = .97$ excluding Lepidoptera), suggesting support for the use of ED as a surrogate for species diversity in survey strategies and biodiversity assessments.

4 | DISCUSSION

4.1 | A novel approach to biological survey programme evaluation

We found that the Bush Blitz programme efficiently contributed large numbers of records of new, or previously unrecorded, species and that these contributed to improving representation of biodiversity in databases, when measured relative to background observations aggregated by the Atlas of Living Australia (ALA). The accrued value of additional investment in targeted survey effort was greatest for those biological groups for which records of occurrence were least well represented across their environmental range. The three invertebrate taxa represented by most new records of species contributed through Bush Blitz surveys (Araneae, Lepidoptera and land snails) were also those for which their ED was least well sampled

ED and new records of species

Relationship between sampling ED and new records of species
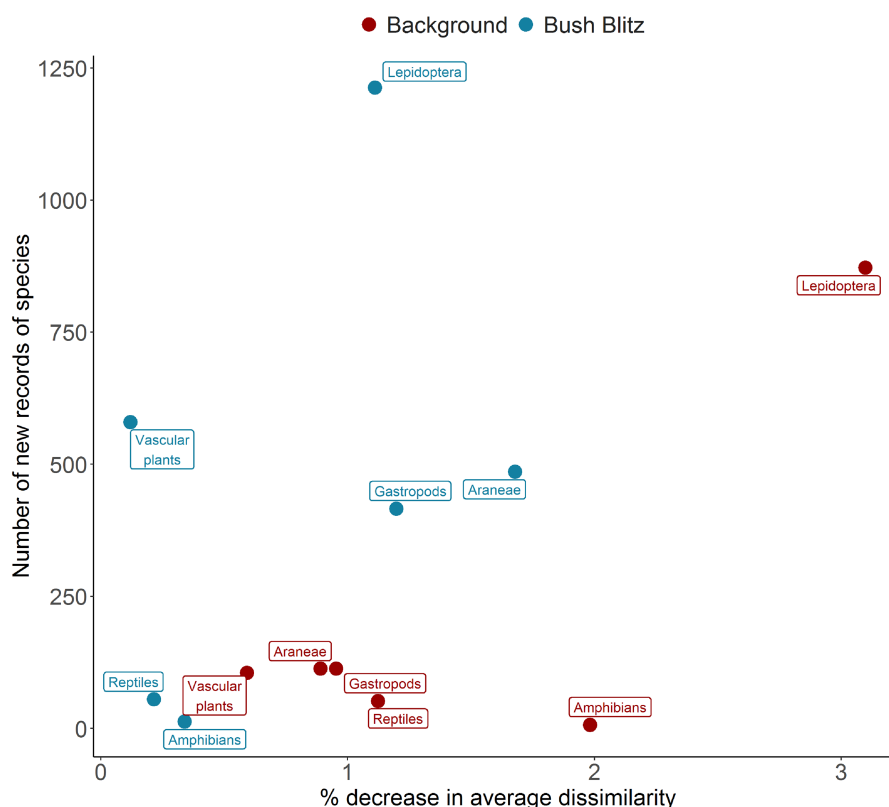
● Background ● Bush Blitz

FIGURE 6 Relationship between gains in sampling environmental diversity (ED) and the proportion of new records of species observed between September 2009 and May 2017, shown for each biological group by Bush Blitz (blue text) or ALA background (red text) records. ED on the *x*-axis in units of percentage decrease in average species compositional dissimilarity between sites without species records to those with records. The *y*-axis is the proportion of new records of species observed relative to the pre-September 2009 baseline. Solid lines represent the linear model fit to each dataset, whereas dashed lines represent the linear model fit to each dataset when the outlier taxa Amphibians and Lepidoptera were removed.

at baseline (Table 1, Figure 5, Figures S3.2–S3.7 in Appendix S3). These results are consistent with findings from similar studies (Albuquerque & Beier, 2018; Engelbrecht et al., 2016; Funk et al., 2005; Guerin et al., 2020, 2021) which support the use of the ED framework (i.e. complementarity of biodiversity surrogates) for prioritising survey locations.

Grantham et al. (2008) evaluated the return on investment from spending different amounts of money on survey data before undertaking a programme of implementing new protected areas. By simulating the selection of new protected areas using different amounts of survey data, they were able to estimate the cost required to retain *Protea* habitat and represent *Protea* species in selected protected areas. They found that the effectiveness of conservation prioritisation increased only minimally after spending as little as 1/25th of what it was estimated to acquire the full dataset. While their assessment clearly underscores the dynamic nature of the conservation problem (Meir et al., 2004), their results may also reflect the fact that the *Protea* survey locations were not prioritised within an ED framework.

4.2 | The Bush Blitz approach to species discovery was effective at filling knowledge gaps

Our assessment illustrates that the number of records is often less important than how well survey locations have been sampled for complementary ED, when used to develop surrogates for biodiversity management and decisions. In our assessment, for all taxa, more sites and species records were represented by ALA background records during the period we evaluated, yet in all cases, Bush Blitz expeditions were responsible for contributing more records of new, or previously unrecorded, species for the given survey effort (Figure 5). Furthermore, the contribution of additional surveys to species discovery is related to sampling of ED (Figure 6). Therefore, by incorporating the iterative filling of gaps in ED in survey design strategies, we can expect higher rates of return in biodiversity data for decision-making for the same effort, when compared with ad hoc surveys.

Ware et al. (2018) showed that Bush Blitz data led to improvements in the effectiveness of modelled biodiversity surrogates across all six taxonomic groups considered in the present work.

Surrogates are assumed to represent spatial patterns in the distribution of biodiversity, and therefore improvements in how well they achieve this should translate into improved biodiversity assessments and prioritisation decisions. Improvements were greatest for Lepidoptera and Araneae, reflected in the present work by the efficiency with which Bush Blitz records sampled ED relative to the other taxonomic groups (Figure 3). Despite vascular plant ED being already well sampled relative to the other groups considered at baseline (second only to reptiles), the modelled plant surrogate developed by Ware et al. (2018) was still improved, both as a within-taxa and cross-taxon surrogate, with the addition of Bush Blitz data. We did not explore how the value of additional data might vary when factoring in financial costs associated with the Bush Blitz expeditions, but these would be partially offset for any given taxonomic group because each Bush Blitz survey targets numerous taxa.

Few of the species accumulation curves representing Bush Blitz expeditions shown in Figure 5 appear to be plateauing (except perhaps, reptiles and amphibians), from which we could assume that further Bush Blitz discovery expeditions will continue to contribute records of new, or previously unrecorded, species. GDM-based cross-taxon surrogates using either gastropod or amphibian records perform best (Ware et al., 2018), and so additional records of these taxa should yield further improvements when applied to modelled surrogates, adding to their value in biodiversity assessments. Additional data for the less well-represented groups will likely strengthen GDM performance and thereby lead to improved surrogacy performance. This is particularly so for Araneae and Lepidoptera, where additional data may enable replacing the vascular plant proxy model used in this assessment.

4.3 | ED-based complementarity is an effective tool for prioritising new survey locations

The Bush Blitz programme site selection method was aided in later years (2014–2018) by maps identifying sites, which would best complement existing sampling of ED for multiple taxonomic groups (Preece et al., 2015). The effect of incorporating these data in site selection is not readily identifiable in our results because their input was balanced alongside a host of other factors related to site selection and survey logistics.

Species accumulation curves representing records of new, or previously unrecorded, species at sites selected by the *p*-median algorithm (Figure 5) were broadly as, or more, efficient than those where site records were tallied chronologically. The substantial contribution made by Bush Blitz expeditions to gains in ED complementarity for the three invertebrate groups, in particular, is shown in Figure 4. The spatial gain is more localised for relatively well-surveyed groups such as vascular plants (Figure 4). Taken together, these results suggest ED complementarity should be a key consideration when selecting prospective survey sites for the purposes of nature discovery and improving biodiversity assessments for more robust conservation decisions. These findings for individual taxa are similar to those reported

in a retrospective evaluation of an Australian continental network of ecosystem monitoring plots (Guerin et al., 2020). In their analysis, Guerin et al. (2020) found sampling of ED to be a good surrogate for ecological representativeness, and the cumulative sampling of environments was strongly related to the sampling of new records of vascular plant species. In our implementation of ED complementarity, we devised a solution (detailed in Appendix S1.1) allowing us to apply the continuous version ED as originally proposed by Faith (1994), but until now was considered computationally intractable.

4.4 | Relationship between ED and species richness

The relationships between sampling of ED and records of new, or previously unrecorded, species across taxa were weak (Figure 6) and affected strongly by outliers. New Bush Blitz records of Lepidoptera were frequently obtained despite being represented by relatively moderate gains in the sampling of ED. In the case of background data, the gain in sampling of ED for amphibians was relatively good but led to few contributions of new records of species. We expect that these outliers simply represent both ends of the diversity spectrum along with a bias in survey effort: Lepidoptera is a highly diverse order, which was represented by relatively few records at baseline, whereas amphibians are a relatively species-poor class well represented by occurrence records at baseline. Additionally, many of the relatively remote regions of Australia targeted by Bush Blitz expeditions tended to have low amphibian diversity (i.e. too hot and dry or too cool).

Aspects of the surrogate modelling approach can also affect the strength of the identified relationship between ED and species richness. First, the spatial resolution of environmental predictors to which our models were fitted is coarse relative to the scale at which taxa use habitat. Second, we used a surrogate model (the plant model) for Lepidoptera as too few Lepidoptera records were available for this model. Third, the choice of candidate predictor variables used in the amphibian model derived from a generic set, whereas some custom predictors to better account for the life cycle and biology of distributions may result in a more effective surrogate. Refinement in these aspects should yield a better representation of biologically scaled ED. All else being equal, where novel environments remain unsampled, the opportunity to collect new records of species will be greater where diversity is greater, and where existing records at baseline are fewer.

4.5 | Incorporating species richness models to enhance priority site selection using ED

A key omission in the procedure we used to calculate ED complementarity of any survey site to the existing set of surveyed sites was a measure of the local species richness (alpha diversity) expected at each of these sites. While the difference in species composition (beta diversity) between sites is critical in assessing complementarity,

local species richness is required in conjunction with GDM-derived compositional dissimilarity to optimally prioritise sites for the species discovery objective. For example, two unsurveyed sites can be equally dissimilar to an existing surveyed site if they are estimated to share no species in common; in these cases, the priority site to survey for the Bush Blitz species discovery approach would be the one with the greatest species richness. Not incorporating measures of species richness likely led to suboptimal site selections in the analyses conducted in this work, and therefore results based on p-median calculations should be viewed as conservative estimates. Methods to formally incorporate species richness in biodiversity surrogate models have been discussed (Faith & Walker, 1996b) and published elsewhere (Albuquerque & Beier, 2018; Arponen et al., 2008) and should be investigated and incorporated in future complementarity-based prioritisations.

4.6 | The value of the Atlas of Living Australia for survey programme evaluation

ALA records available over the period of our evaluation were dominated by records based on specimens collected by institutes (Figures S3.1a and S3.1b). The end of our study period approximately coincides with a period of exponential growth in the number of Australian observations contributed to the iNaturalist biodiversity citizen science platform (Mesaglio & Callaghan, 2021). The number of species observations contributed increased by nearly an order of magnitude between 2016 and 2021 (Mesaglio & Callaghan, 2021), all of which have since become available in the ALA. Understanding how these different trends in data collection and mobilisation might sway the results we observed would require additional analysis. Large numbers of additional records contributed for certain taxa, for example Lepidoptera (Mesaglio & Callaghan, 2021), would almost certainly include records of new, or previously unrecorded, species. How efficiently this sampling contributed records of new, or previously unrecorded, species, however, would depend in large part on how well additional records sampled new ED and complemented existing biases in the ALA. Whether there are large-scale differences between what was predominately institution-provided data and the newly aggregated citizen science data in these measures seems likely, but it is difficult to speculate on how these differences might alter the trends reported here without further analysis.

Taxonomic and spatial observation bias in the ALA is strong (Daru et al., 2018; Haque et al., 2017, 2020), as in other large public repositories of biodiversity data (Daru et al., 2018; Troudet et al., 2017). These biases, due to ad hoc or opportunistic data aggregation (including repeat sampling for monitoring purposes), necessarily reduce the relative efficiency of spatial sampling of ED or records of new, or previously unrecorded, species. Furthermore, large collections by universities or private institutes may not (yet) be mobilised through the ALA (Nelson & Ellis, 2019). We are aware of datasets for particular taxa including Lepidoptera and reptiles that were not available

through the ALA at the time we extracted data, and their omission likely affects our results. As with the consideration of additional citizen science records outlined above, it is uncertain how these missing data might alter both the sampling of ED and efficiency of generating records of new, or previously undescribed, species that we observed. In this sense, our use of the ALA as a benchmark conflates the true state of species discovery with the state of species occurrence record publishing, and therefore, our assessment of Bush Blitz programme efficiency for some taxa may be inflated. Nevertheless, the ALA represents the leading Australian platform in aggregating and providing publicly accessible biodiversity information, with the aim of enabling evidence-based decision-making in all aspects of biodiversity and environmental research and policy (ALA, 2020; Belbin et al., 2021). Future conservation decisions based on best-available data, including their use in surrogate models, will depend on how comprehensively the ALA represents biodiversity. Part of the value of the Bush Blitz programme is the commitment obtained from partner natural history institutions to prioritise digital mobilisation of all collected data into the ALA for public access.

5 | CONCLUSION

Our evaluation highlights the value of targeted contributions made by Bush Blitz relative to background species observations reflected in the ALA and demonstrates the effectiveness with which the programme sampled ED which translated into records of new, or previously unrecorded, species. The retrospective nature of our analysis and use of the ALA as the source of biological heritage data on species distributions is timely, given Australia's achievements since 2010 in mobilising vast collections of natural history museum specimen and ecological survey data to publicly accessible repositories (Belbin et al., 2021; Belbin & Williams, 2016; Nelson & Ellis, 2019; Sparrow et al., 2020; Turner et al., 2017). Our assessment supports the use of the ED framework in guiding future survey priorities and assessing existing survey adequacy and underscores the need to continue to mobilise survey data into publicly accessible repositories. It further provides a framework to inform data aggregation strategies, which might be adopted by data aggregators, such as the ALA, to incentivise and prioritise data digitisation. While background efforts contributed large numbers of records to our collective knowledge of biodiversity, our assessment indicates that targeted, data-driven surveys are required to efficiently enhance data underpinning capacity to make better-informed biodiversity assessments and decisions.

ALA. We are also grateful to two anonymous reviewers and the editor for their suggestions, which improved an earlier version of this manuscript.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ddi.13806.

DATA AVAILABILITY STATEMENT

Biological occurrence records underpinning the findings in this study are publicly available from the Atlas of Living Australia (https://www.ala.org.au/) and via the link https://doi.org/10.5061/dryad.866t1g1wr. The four generalised dissimilarity models used as surrogates for ED are available for download via www.data.csiro.au: vascular plants (Williams et al., 2013); reptiles (Williams et al., 2014c); amphibians (Williams et al., 2014a); and land snails (Williams et al., 2014b). Demand points generated for the analyses of ED complementarity, and maps of improvements in environmental diversity coverage (described in Section 2.3 and the Appendix) are available via the link https://doi.org/10.5061/dryad.866t1g1wr.

CONFLICT OF INTEREST STATEMENT

The authors have nothing to disclose. The analyses reported here were supported with funding through Australia's Bush Blitz programme to CSIRO.

ORCID

Chris Ware 🄳 https://orcid.org/0000-0001-6739-1331

REFERENCES

ABRS. (2009). *Charles Darwin reserve WA (3–9 May, 14–25 September, 7–18 December 2009) and Kadji Kadji, Karara, Lochada reserves WA (14–25 September, 7–18 December 2009). A Bush Blitz Survey Report*. Australian Biological Resources Study (ABRS), Commonwealth of Australia.

ABRS. (2017a). *Bradshaw field training area Northern Territory, 8–19 May 2017. A Bush Blitz Survey Report*. Australian Biological Resources Study (ABRS), Commonwealth of Australia.

ABRS. (2017b). *Carnarvon Station Reserve Queensland, 7–17 October 2014. A Bush Blitz Survey Report*. Australian Biological Resources Study (ABRS), Commonwealth of Australia.

ALA. (2020). *Atlas of living Australia ala.org.au: Strategy 2020–2025*. CSIRO.

ALA. (2023). *Data on the ALA: ALA support*. Atlas of Living Austraalia. https://support.ala.org.au/support/solutions/articles/6000261411-data-on-the-ala

Albuquerque, F., & Beier, P. (2018). Improving the use of environmental diversity as a surrogate for species representation. *Ecology and Evolution*, 8, 852–858.

Allnutt, T. F., Ferrier, S., Manion, G., Powell, G. V. N., Ricketts, T. H., Fisher, B. L., Harper, G. J., Irwin, M. E., Kremen, C., Labat, J. N., Lees, D. C., Pearce, T. A., & Rakotondrainibe, F. (2008). A method for quantifying biodiversity loss and its application to a 50-year record of deforestation across Madagascar. *Conservation Letters*, 1, 173–181.

Arponen, A., Moilanen, A., & Ferrier, S. (2008). A successful community-level strategy for conservation prioritization. *Journal of Applied Ecology*, 45, 1436–1445.

Ashcroft, M. B., Gollan, J. R., Faith, D. P., Carter, G. A., Lassau, S. A., Ginn, S. G., Bulbert, M. W., & Cassis, G. (2010). Using generalised dissimilarity models and many small samples to improve the efficiency of regional and landscape scale invertebrate sampling. *Ecological Informatics*, 5, 124–132.

Austin, A. D., Yeates, D. K., Cassis, G., Fletcher, M. J., la Salle, J., Lawrence, J. F., McQuillan, P. B., Mound, L. A., J Bickel, D., Gullan, P. J., Hales, D. F., & Taylor, G. S. (2004). Insects 'down under' – Diversity, endemism and evolution of the Australian insect fauna: Examples from select orders. *Australian Journal of Entomology*, 43, 216–234.

Balmford, A., & Gaston, K. J. (1999). Why biodiversity surveys are good value. *Nature*, 398, 204–205.

Beier, P., & de Albuquerque, F. S. (2016). Evaluating β diversity as a surrogate for species representation at fine scale. *PLoS ONE*, 11, e0151048.

Beier, P., & de Albuquerque, F. S. J. C. B. (2015). Environmental diversity as a surrogate for species representation. *Conservation Biology*, 29, 1401–1410.

Belbin, L., Chapman, A., Wieczorek, J., Morris, P. J., & Zermoglio, P. (2020). Task group 2 – Data quality tests and assertions. *Biodiversity Information Science and Standards*, 4, e58982.

Belbin, L., Wallis, E., Hobern, D., & Zerger, A. (2021). The atlas of living Australia: History, current state and future directions. *Biodiversity Data Journal*, 9, e65023.

Belbin, L., & Williams, K. J. (2016). Toward a national bio-environmental data facility: Experiences from the atlas of living Australia. *International Journal of Geographic Information Sciences*, 30, 108–125.

Bell, K. L., Heard, T. A., Manion, G., Ferrier, S., & van Klinken, R. D. (2014). Characterising the phytophagous arthropod fauna of a single host plant species: Assessing survey completeness at continental and local scales. *Biodiversity and Conservation*, 23, 2985–3003.

Brito, D. (2010). Overcoming the Linnean shortfall: Data deficiency and biological survey priorities. *Basic and Applied Ecology*, 11, 709–713.

Callaghan, C. T., Poore, A. G., Major, R. E., Rowley, J. J. L., & Cornwell, W. K. (2019). Optimizing future biodiversity sampling by citizen scientists. *Proceedings of the Royal Society. B. Biological sciences*, 286(1912), 20191487.

Carvalho, S. B., Brito, J. C., Pressey, R. L., Crespo, E., & Possingham, H. P. (2010). Simulating the effects of using different types of species distribution data in reserve selection. *Biological Conservation*, 143, 426–438.

Chapman, A. D. (2009). *Numbers of living species in Australia and the world* (2nd ed.). Australian Biological Resources Study: Australian Government.

Cheng, M., & Cassis, G. (2019). Systematic study of the Australian plant bug genus *Xasmasoma*, gen. Nov.(Insecta: Heteroptera: Miridae: Orthotylinae), including host plant and biogeographic analysis, and description of twelve new species. *Invertebrate Systematics*, 33, 1–70.

Chowdhury, S., Fuller, R. A., Ahmed, S., Alam, S., Callaghan, C. T., Das, P., Correia, R. A., Di Marco, M., Di Minin, E., Jarić, I., Labi, M. M., Ladle, R. J., Rokonuzzaman, M., Roll, U., Sbragaglia, V., Siddika, A., & Bonn, A. (2023). Using social media records to inform conservation planning. *Conservation Biology*, e14161. https://doi.org/10.1111/cobi.14161

Church, R. L. (2002). Geographical information systems and location science. *Computers & Operations Research*, 29, 541–562.

Coetzee, B. W. T., Gaston, K. J., & Chown, S. L. (2014). Local scale comparisons of biodiversity as a test for global protected area ecological performance: A meta-analysis. *PLoS ONE*, 9, e105824.

Daru, B. H., Park, D. S., Primack, R. B., Willis, C. G., Barrington, D. S., Whitfeld, T. J. S., Seidler, T. G., Sweeney, P. W., Foster, D. R., Ellison, A. M., & Davis, C. C. (2018). Widespread sampling biases

in herbaria revealed from large-scale digitization. *New Phytologist*, *217*, 939–955.

DAWE. (2021). *Collaborative Australian protected areas database (CAPAD) 2020 – Terrestrial*. Australian Government Department of Agriculture, Water and the Environment.

Di Marco, M., Harwood, T. D., Hoskins, A. J., Ware, C., Hill, S. L. L., & Ferrier, S. (2019). Projecting impacts of global climate and land-use scenarios on plant biodiversity using compositional-turnover modelling. *Global Change Biology*, *25*, 2763–2778.

Dunn, J. C., Buchanan, G. M., Stein, R. W., Whittingham, M. J., & McGowan, P. J. K. (2016). Optimising different types of biodiversity coverage of protected areas with a case study using Himalayan Galliformes. *Biological Conservation*, *196*, 22–30.

Engelbrecht, I., Robertson, M., Stoltz, M., & Joubert, J. W. (2016). Reconsidering environmental diversity (ED) as a biodiversity surrogacy strategy. *Biological Conservation*, *197*, 171–179.

Faith, D. P. (1994). Phylogenetic pattern and the quantification of organismal biodiversity. *Philosophical Transactions of the Royal Society of London Series B Biological Sciences*, *345*, 45–58.

Faith, D. P. (2003). Environmental diversity (ED) as surrogate information for species-level biodiversity. *Ecography*, *26*, 374–379.

Faith, D. P., Ferrier, S., & Walker, P. A. (2004). The ED strategy: How species-level surrogates indicate general biodiversity patterns through an 'environmental diversity' perspective. *Journal of Biogeography*, *31*, 1207–1217.

Faith, D. P., & Walker, P. A. (1996a). 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.

Faith, D. P., & Walker, P. A. (1996b). How do indicator groups provide information about the relative biodiversity of different sets of areas?: On hotspots, complementarity and pattern-based approaches. *Biodiversity Letters*, *3*, 18–25.

Ferrier, S. (2002). Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? *Systematic Biology*, *51*, 331–363.

Ferrier, S., Drielsma, M., Manion, G., & Watson, G. (2002). Extended statistical approaches to modelling spatial pattern in biodiversity in Northeast New South Wales. II. Community-level modelling. *Biodiversity and Conservation*, *11*, 2309–2338.

Ferrier, S., Harwood, T. D., Ware, C., & Hoskins, A. J. (2020). A globally applicable indicator of the capacity of terrestrial ecosystems to retain biological diversity under climate change: The bioclimatic ecosystem resilience index. *Ecological Indicators*, *117*, 106554.

Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, *13*, 252–264.

Ferrier, S., Powell, G. V. N., Richardson, K. S., Manion, G., Overton, J. M., Allnutt, T. F., Cameron, S. E., Mantle, K., Burgess, N. D., Faith, D., Lamoreux, J., Kier, G., Hijmans, R., Funk, V. A., Cassis, G., Fisher, B. L., Flemons, P., Lees, D. C., Lovett, J. C., & Van Rompaey, R. (2004). Mapping more of terrestrial biodiversity for global conservation assessment. *Bioscience*, *54*, 1101–1109.

Field, C. R., & Elphick, C. S. (2019). Quantifying the return on investment of social and ecological data for conservation planning. *Environmental Research Letters*, *14*(12), 124081.

Funk, V. A., Richardson, K. S., & Ferrier, S. (2005). Survey-gap analysis in expeditionary research: Where do we go from here? *Biological Journal of the Linnean Society*, *85*, 549–567. https://doi.org/10.1111/j.1095-8312.2005.00520.x

Gaston, K. J., & Rodrigues, A. S. L. (2003). Reserve selection in regions with poor biological data. *Conservation Biology*, *17*, 188–195.

GBIF Secretariat. (2019). *GBIF science review 2019*. Global Biodiversity Information Facility.

Grantham, H. S., Moilanen, A., Wilson, K. A., Pressey, R. L., Rebelo, T. G., & Possingham, H. P. (2008). Diminishing return on investment for biodiversity data in conservation planning. *Conservation Letters*, *1*, 190–198.

Guerin, G. R., Williams, K. J., Leitch, E., Lowe, A. J., & Sparrow, B. (2021). Using generalised dissimilarity modelling and targeted field surveys to gap-fill an ecosystem surveillance network. *Journal of Applied Ecology*, *58*, 766–776.

Guerin, G. R., Williams, K. J., Sparrow, B., & Lowe, A. J. (2020). Stocktaking the environmental coverage of a continental ecosystem observation network. *Ecosphere*, *11*, e03307.

Hanson, J. O., McCune, J. L., Chadès, I., Proctor, C. A., Hudgins, E. J., & Bennett, J. R. (2023). Optimizing ecological surveys for conservation. *Journal of Applied Ecology*, *60*(1), 41–51.

Haque, M. M., Beaumont, L. J., & Nipperess, D. A. (2020). Taxonomic shortfalls in digitised collections of Australia's flora. *Biodiversity and Conservation*, *29*, 333–343.

Haque, M. M., Nipperess, D. A., Gallagher, R. V., & Beaumont, L. J. (2017). How well documented is Australia's flora? Understanding spatial bias in vouchered plant specimens. *Austral Ecology*, *42*, 690–699.

Kujala, H., Lahoz-Monfort, J. J., Elith, J., & Moilanen, A. (2018). Not all data are equal: Influence of data type and amount in spatial conservation prioritisation. *Methods in Ecology and Evolution*, *9*(11), 2249–2261.

Lõhmus, A., Lõhmus, P., & Runnel, K. (2018). A simple survey protocol for assessing terrestrial biodiversity in a broad range of ecosystems. *PLoS ONE*, *13*, e0208535.

Maxwell, S. L., Rhodes, J. R., Runge, M. C., Possingham, H. P., Ng, C. F., & McDonald-Madden, E. (2015). How much is new information worth? Evaluating the financial benefit of resolving management uncertainty. *Journal of Applied Ecology*, *52*(1), 12–20.

Meir, E., Andelman, S., & Possingham, H. P. (2004). Does conservation planning matter in a dynamic and uncertain world? *Ecology Letters*, *7*, 615–622.

Mentges, A., Blowes, S. A., Hodapp, D., Hillebrand, H., & Chase, J. M. (2021). Effects of site-selection bias on estimates of biodiversity change. *Conservation Biology*, *35*(2), 688–698.

Mesaglio, T., Callaghan, C., Samonte, F., Gorta, S. B. Z., & Cornwell, W. K. (2023). Recognition and completeness: Two key metrics for judging the utility of citizen science data. *Frontiers in Ecology and the Environment*, *21*(4), 167–174.

Mesaglio, T., & Callaghan, C. T. (2021). An overview of the history, current contributions and future outlook of iNaturalist in Australia. *Wildlife Research*, *48*(4), 289–303.

Miles, N. (2011). *Guide to data quality, version 1.2* (p. 17). Atlas of Living Australia.

Mokany, K., Ferrier, S., Harwood, T. D., Ware, C., di Marco, M., Grantham, H. S., Venter, O., Hoskins, A. J., & Watson, J. E. M. (2020). Reconciling global priorities for conserving biodiversity habitat. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 9906–9911.

National Reserve System Task Group. (2009). *Australia's strategy for the National Reserve System 2009–2030*. Department of the Environment, Water, Heritage and the Arts, Australian Government.

Nelson, G., & Ellis, S. (2019). The history and impact of digitization and digital data mobilization on biodiversity research. *Philosophical Transactions of the Royal Society B*, *374*, 20170391.

Newman, P., Raymond, B., et al. (2019). *ALA4R version 1.7.0*. Atlas of Living Australia. https://rforge.net/ALA4R/index.html

Oliver, R. Y., Meyer, C., Ranipeta, A., Winner, K., & Jetz, W. (2021). Global and national trends, gaps, and opportunities in documenting and monitoring species distributions. *PLoS Biology*, *19*, e3001336.

Possingham, H. P., Grantham, H., & Rondinini, C. (2007). How can you conserve species that haven't been found? *Journal of Biogeography*, *34*, 758–759.

Preece, M., Harding, J., & West, J. G. (2015). Bush blitz: Journeys of discovery in the Australian outback. *Australian Systematic Botany, 27,* 325–332.

R Core Team. (2020). *R: A language and environment for statistical computing. Version 3.4.4.* R Foundation for Statistical Computing. https://www.R-project.org

Raymond, B., VanDerWal, J., et al. (2014). *ALA4R version 1.01.* Atlas of Living Australia. https://rforge.net/ALA4R/index.html

Reddy, C. S. (2010). Survey-gap analysis for botanical research using integrated approach through taxonomical data and geoinformatics. *Journal of the Indian Society of Remote Sensing, 38,* 577–584.

Reside, A. E., VanDerWal, J., et al. (2013). *Climate change refugia for terrestrial biodiversity.* National Climate Change Adaptation Research Facility.

ReVelle, C. S., & Swain, R. W. (1970). Central facilities location. *Geographical Analysis, 2,* 30–42.

Rodewald, A. D., Strimas-Mackey, M., Schuster, R., & Arcese, P. (2019). Tradeoffs in the value of biodiversity feature and cost data in conservation prioritization. *Scientific Reports, 9,* 15921.

Sparrow, B. D., Foulkes, J. N., Wardle, G. M., Leitch, E. J., Caddy-Retalic, S., van Leeuwen, S. J., Tokmakoff, A., Thurgate, N. Y., Guerin, G. R., & Lowe, A. J. (2020). A vegetation and soil survey method for surveillance monitoring of rangeland environments. *Frontiers in Ecology and Evolution, 8,* 1–18.

Taxonomy Decadal Plan Working Group. (2018). *Discovering diversity: A decadal plan for taxonomy and biosystematics in Australia and New Zealand 2018–2028.* Australian Academy of Science and Royal Society Te Apārangi.

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports, 7,* 9132.

Turner, D. J., Smyth, A. K., et al. (2017). ÆKOS – Next generation online data and information infrastructure for the ecological science community. In A. Chabbi & H. W. Loescher (Eds.), *Terrestrial ecosystem research infrastructures: Challenges and opportunities* (pp. 341–533). CRC Press (Taylor & Francis Group), Ækos.

Ware, C., Williams, K. J., Harding, J., Hawkins, B., Harwood, T., Manion, G., Perkins, G. C., & Ferrier, S. (2018). Improving biodiversity surrogates for conservation assessment: A test of methods and the value of targeted biological surveys. *Diversity and Distributions, 24,* 1333–1346.

Williams, K., Manion, G., et al. (2014a). *AMP_V2_R2: Generalised dissimilarity model of compositional turnover in amphibian species for continental Australia at 9 second resolution using ALA data extracted February 2014. v1.* (ed. by CSIRO Land and Water). CSIRO. Data Collection. https://doi.org/10.4225/08/586ef70d640f4

Williams, K., Manion, G., et al. (2014b). *GAS_V3_R2: Generalised dissimilarity model of compositional turnover in land snail species for continental Australia at 9 second resolution using ANHAT data extracted April 2013.* (ed. by CSIRO Land and Water). CSIRO. Data Collection. https://doi.org/10.4225/08/586ec6e4246ac

Williams, K., Manion, G., et al. (2014c). *REP_V2_R3: Generalised dissimilarity model of compositional turnover in reptile species for continental Australia at 9 second resolution using ALA data extracted February

2014. v1.* (ed. by CSIRO Land and Water). CSIRO. Data Collection. https://doi.org/10.4225/08/586ec70f080bf

Williams, K. J., Belbin, L., Austin, M. P., Stein, J. L., & Ferrier, S. (2012). Which environmental variables should I use in my biodiversity model? *International Journal of Geographic Information Sciences, 26,* 2009–2047.

Williams, K. J., Harwood, T., et al. (2013). *VAS_v5_r11: Generalised dissimilarity model of compositional turnover in vascular plant species for continental Australia at 9 second resolution using ANHAT data extracted April 2013 (v1).* (ed. by CSIRO Land and Water). CSIRO. Data Collection. https://doi.org/10.4225/08/557FB520465F7

Williams, K. J., Harwood, T. D., et al. (2016). *Assessing the ecological representativeness of Australia's terrestrial National Reserve System: A community-level modelling approach (a report prepared for the Australian government, Department of the Environment).* CSIRO Land and Water (EP163634).

Zborowski, P., & Edwards, T. (2007). *A guide to Australian moths.* CSIRO Publishing.

BIOSKETCH

Collectively, the author team's research interests are in conservation assessment and macroecology. By documenting plants and animals across Australia and developing approaches to biodiversity modelling, the team's research aims are to improve our understanding of current patterns in biodiversity and our capacity to project future outcomes for biodiversity under scenarios of change.

Author contributions: S Ferrier, KJW/CW, GM, DPF and JH conceived and designed the project; CW/KJW, JP, BH, JH, TDH, AL and JM acquired the data; KJW/CW, GM and S Ferrier analysed and interpreted the data; KJW/CW, S Ferrier, DPF, JH and S Fyfe wrote the paper.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.