Vol. 688: 167–172, 2022
https://doi.org/10.3354/meps14027

MARINE ECOLOGY PROGRESS SERIES
Mar Ecol Prog Ser

Published April 28



NOTE

# Using the background of fish photographs to quantify habitat composition in marine ecosystems

**Madison H. Bolt[1,4,*], Corey T. Callaghan[2,3], Alistair G. B. Poore[1,2], Adriana Vergés[1,2], Christopher J. Roberts[1,2]**

[1]Centre for Marine Science and Innovation, School of Biological, Earth and Environmental Sciences, UNSW Sydney, Sydney 2052, NSW, Australia

[2]Ecology & Evolution Research Centre, School of Biological, Earth and Environmental Sciences, UNSW Sydney, Sydney 2052, NSW, Australia

[3]German Centre for Integrative Biodiversity Research (iDiv), 04103 Leipzig, Germany

[4]*Present address:* Department of Botany and Zoology, UBC Vancouver, Vancouver, BC V6T1Z4, Canada

ABSTRACT: Citizen science initiatives that collect opportunistic photos, or recordings, of living organisms (e.g. iNaturalist) are increasingly recognized for their importance in monitoring biodiversity. These projects are focussed primarily on recording the occurrence of individual species in space and time. Each photo potentially also contains additional valuable information. Here, we explored the amount and potential value of background information captured in fish photographs as a method to characterise reef habitats. The habitat in the background of fish photographs shared on iNaturalist was analysed for 6 sites across Australia. To measure accuracy of the habitat data captured in the iNaturalist photos, the habitat composition of each site was compared to standardised photo-quadrats from the citizen science project Reef Life Survey (RLS). Across all sites, 70–85% of the fish photographs from iNaturalist contained discernible biotic habitat in the background. Habitat composition as measured from the background of opportunistic fish photographs was similar to those of standardised surveys from RLS. In the face of rapid environmental change, opportunistic photographs collected by recreational divers represent a complementary way to rapidly and cost-effectively collect habitat data at more reefs and more frequently than is generally feasible with standardised scientific surveys.

KEY WORDS: Citizen science · Community science · iNaturalist · Opportunistic data · Biodiversity · Temperate reefs · Coral reefs

## 1. INTRODUCTION

The quantity and diversity of citizen science projects, also referred to as community science or contributory science, has increased dramatically in recent decades with an associated growth in use for monitoring biodiversity (Pocock et al. 2017, McKinley et al. 2017). Citizen science can provide a cost-effective supplement or alternative to often expensive and time-consuming data collection by professional scientists (Thornhill et al. 2016, Poisson et al. 2020).

*Corresponding author: maddybolt@alumni.ubc.ca



Publisher: Inter-Research · www.int-res.com

iNaturalist (www.inaturalist.org) is a successful citizen science project with more than 89 million opportunistic observations of over 344 000 species (as of February 2022). iNaturalist data have been used for vegetation mapping (Uyeda et al. 2020), monitoring urban biodiversity (Callaghan et al. 2020), detecting range extensions of alien species (Agarwal 2017) and the rediscovery of 'lost' species (Richart et al. 2019). These advances are largely focussed on species occurrences, but each photo potentially contains additional ecological information including interspecific interactions, phenotypic traits, breeding status and habitat associations (Callaghan et al. 2021).

Due to their charismatic nature, fish are often the primary subject of many underwater photos (Troudet et al. 2017). However, important 'incidental' habitat information is often captured in the background of these photographs (Fig. 1a,b), reflecting a possible alternative, complementary, method to identify major habitat-forming organisms such as macroalgae, seagrasses and corals. Here, we investigated whether iNaturalist fish photographs could contain valuable additional data for cost-effective monitoring of reef habitats.

To demonstrate the potential value of incidental habitat data available in citizen science photographs, we quantified the proportion of iNaturalist fish photographs with identifiable benthic habitats in the background for 6 sites from temperate to tropical Australia. To assess the accuracy of iNaturalist in determining broadscale habitat composition, we compared the presence/absence of several habitat forming benthic organisms (macroalgae, sponges etc.) between iNaturalist photographs and standardised photo-quadrats from Reef Life Survey (RLS; https://reeflifesurvey.com) (Edgar & Stuart-Smith 2014).

## 2. MATERIALS AND METHODS

### 2.1. Data sources

iNaturalist is a citizen science platform for participants to share opportunistic observations of any

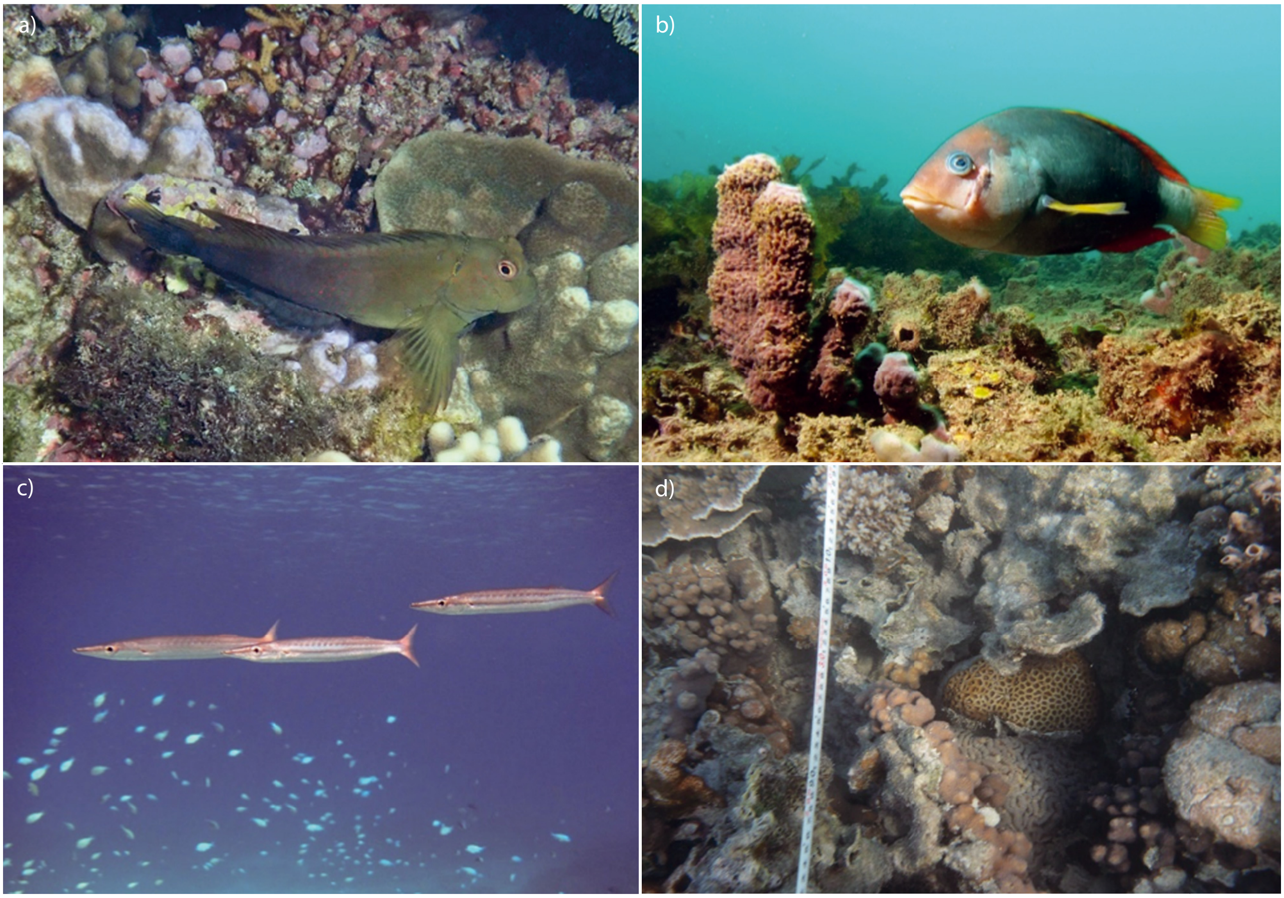


Fig. 1. Example photographs from iNaturalist and Reef Life Survey (RLS). (a,b) iNaturalist fish photographs with usable habitat information in background. (c) iNaturalist image with no information on benthic habitat in the background. (d) RLS standardised photo-quadrat of benthic habitat. Photo credits: (a) Jeyre, (b) John Turnbull, (c) Geoff Shuetrim

organisms, which are then identified to the lowest possible taxonomic resolution by iNaturalist users. RLS is a citizen science initiative which trains volunteer divers to conduct standardised scientific surveys, and all survey data are made publicly available (Edgar & Stuart-Smith 2014). RLS monitors the benthic habitat by taking 20 photographs of approximately 0.3 × 0.3 m of seabed (i.e. photo-quadrats) along a 50 m transect.

The 6 study sites in Australia (Fig. S1 in the Supplement at www.int-res.com/articles/suppl/m688p167_supp.pdf) were chosen based on having at least 200 iNaturalist fish observations from a 5 $km^2$ area and data from at least 4 RLS surveys. Fifty random fish photographs from each site were downloaded on 21 June 2019 from the iNaturalist project Australasian Fishes (www.inaturalist.org/projects/australasian-fishes). Four RLS transects from between 2015 and 2019 were randomly selected for each site, and 15 images were randomly selected for each survey. The 60 photo-quadrats for each site were downloaded on 19 December 2019 from the RLS database (https://reeflifesurvey.com/survey-data/).

### 2.2. Image classification

The background of each iNaturalist image was first classified as 'usable' or as 'unusable', based on whether habitat-forming organisms could be distinguished in the background. The background of unusable images was further categorised as (1) blurry, (2) dark, (3) sand only, (4) water only, or (5) the subject only (i.e. the fish filled the whole photo) (Fig. S2). For 'usable' images, the background was scored for the presence/absence of the following biotic habitats: turf algae, encrusting algae, macroalgae, seagrass, coral, soft coral and sponge/ascidian (Fig. S3). These broad taxonomic groupings were chosen as they were likely to be distinguishable in the background of both close-up and wide-angle photographs. The image classification was done in the software package photoQuad version 1.4 (Trygonis & Sini 2012).

The RLS photo-quadrats were scored for the presence/absence of each habitat using the same method as the iNaturalist images. The presence/absence of each habitat type was used instead of the more conventional measure of percentage cover within photo-quadrats due to the highly variable area of habitat captured in fish photographs making percentage cover an inconsistent measure (Fig. 1a,b). We then calculated the relative occurrence of each habitat type (how often each habitat was seen relative to all habitats combined), as a measure of how common each habitat type was at a site-wide scale.

### 2.3. Analyses

The relative frequency of each habitat type recorded in the background of iNaturalist photographs was contrasted to RLS photo-quadrats using a linear model, with habitat types pooled, run in the R package 'emmeans' (Lenth 2021). The 95% confidence interval for the slope of the linear model was obtained using the 'confint' function to test for a 1:1 relationship.

To test how robust the relationships between iNaturalist and RLS were to sampling effort (i.e. number of photographs included), the photographs were resampled 1000 times for random subsets of 15, 20, 25 and 30 photographs, and the $R^2$ values from the linear models were recalculated for each run.

A second linear model was run including the interaction between site and habitat to test whether the relationship between iNaturalist and RLS was consistent across the replicate sites regardless of habitat types. The confidence interval for each site was extracted using the 'lstrends' function of 'emmeans' to test for a 1:1 relationship between iNaturalist and RLS at each site.

## 3. RESULTS

Habitat-forming organisms could be identified in the background of between 68 and 86% of iNaturalist fish photographs (Fig. 2). The main reasons for photographs not having usable biotic habitat data were that the background contained sand only (6.4% of photos on average) or water only (5%), or the background was out of focus (6%).

Within the usable iNaturalist photographs, the relative occurrence of major habitat types per site was strongly correlated to the standardised RLS photo-quadrats at the same sites ($p < 0.001$, $R^2 = 0.71$). In addition, the relative occurrence of the different habitat types had a slope close to 1 ($\beta = 0.81 \pm 0.17$ 95% CI), indicating only a slight deviation from a 1:1 relationship between RLS and iNaturalist for all sites combined (Fig. 3a). Resampling of the iNaturalist photographs showed that the positive relationship between data sources was relatively robust, with similar $R^2$ values obtained when the number of photographs was reduced to both 30 and 25 (Fig. S4).

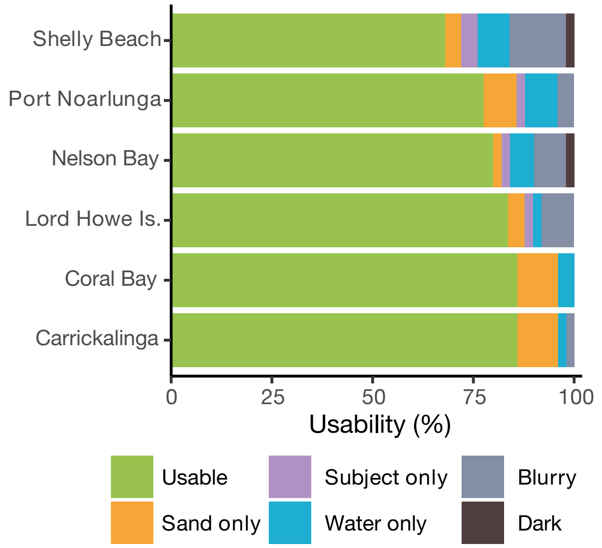


Fig. 2. Percentage of fish photographs from 6 popular recreational dive sites that contained usable information on the composition of habitat-forming organisms captured in the background. The reasons for photographs not having usable biotic habitat information in the background are also shown

Some habitat types appeared to be consistently over- or under-represented by iNaturalist. For example, ascidians and sponges were on average recorded 6.9% more frequently by iNaturalist than RLS, while turf algae were recorded 5% more (Fig. 3). In contrast, macroalgae were recorded 6.4% less frequently by iNaturalist than RLS, while soft corals were recorded 3.3% less and encrusting algae 3.0% less.

When site was included in the model, the interaction term was not significant ($F_{5,30}$ = 0.795, p = 0.56), indicating that the habitat occurrence frequency relationship between iNaturalist and RLS was consistent among study sites (Fig. 3b). The relationship between iNaturalist and RLS was close to 1:1 at most of the study sites, with slopes ranging between 0.779 and 1.025, with confidence intervals overlapping a slope of 1. The only exception was Carrickalinga, which had a slope of 0.46 and a confidence interval of between 0.004 and 0.915.

## 4. DISCUSSION

Citizen science databases are continually increasing (Pocock et al. 2017); for example, iNaturalist alone averaged 68 000 observations per day in 2020. Maximizing the information extracted from this

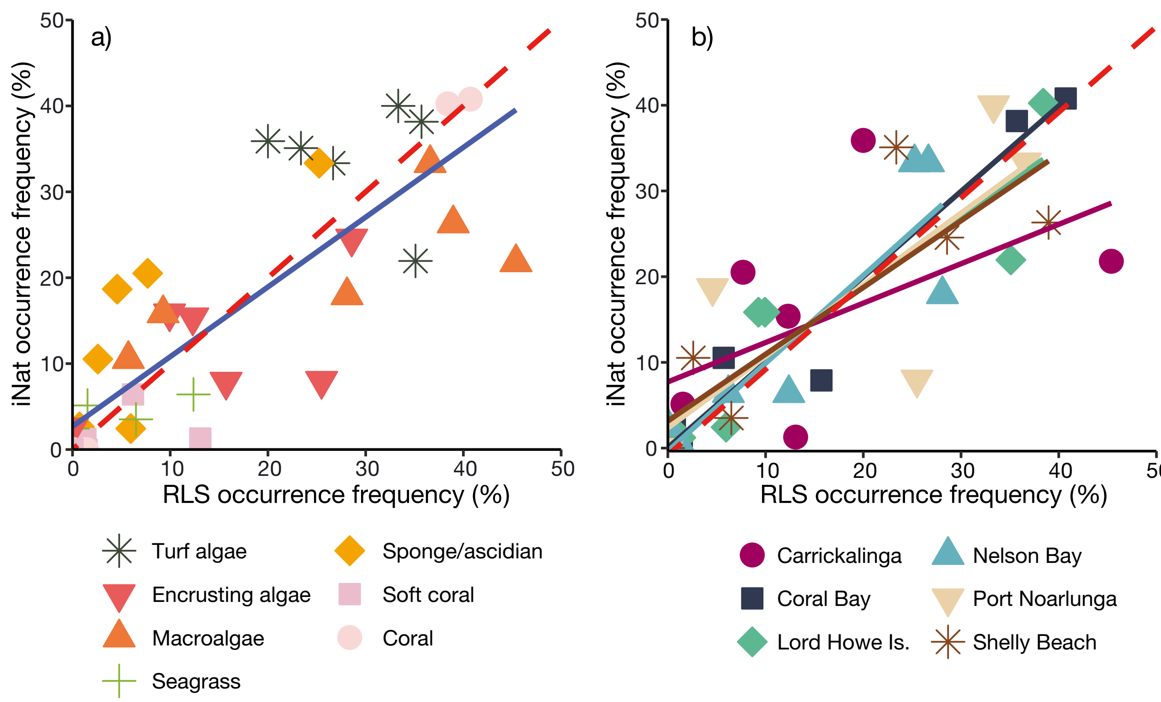


Fig. 3. Relative occurrence of each biotic habitat category recorded by iNaturalist (iNat) and Reef Life Survey (RLS). (a) Overall linear relationship between iNaturalist and RLS (blue solid line) pooling sites and (b) the relationship for each site (coloured solid lines). The red (dashed) line is a 1:1 relationship as expected if both methods recorded an equal frequency of habitat. Points above (below) the dashed line indicate habitats more frequently recorded by iNaturalist (RLS)

resource may assist in timely environmental monitoring and associated management actions. Here, we demonstrated that a high proportion of fish photographs sampled from iNaturalist contain background information that can be used to categorise reef habitats. We also found that the habitat captured in the background of fish photographs was similar to that recorded by standardised habitat surveys, demonstrating the potential utility of 'incidental' habitat data.

Scientific monitoring of marine habitats is limited considerably by the cost and logistical difficulties of doing underwater surveys. In contrast, recreational divers collectively dive more regularly than professional scientists and visit more sites. As such, the use of information in the background of fish photographs could greatly increase marine habitat data availability both spatially and temporally. This is demonstrated by the large numbers and broad spatial coverage of images on iNaturalist, with over 860 000 fish photographs from all around the world (as of February 2022). If a substantial portion of these photographs contain useful habitat data, as demonstrated by this study, this is a considerable amount of habitat information which is not currently being used. However, before such techniques can be implemented into marine monitoring, larger-scale comparisons would be needed to thoroughly test the accuracy and reliability of using underwater photographs to assess habitat.

To date, there have been many comparisons showing that trained citizen scientists can generate comparable data to professionals when using standardised methods (Aceves-Bueno et al. 2017). Although comparisons between opportunistic observations and standardised surveys are limited, some studies have shown correlations between the abundances recorded by these approaches (Snäll et al. 2011, Kamp et al. 2016). However, some discrepancies have also been noted, such as observer biases toward photogenic species (Prudic et al. 2018) or common species not being regularly reported (Snäll et al. 2011). By focussing on the background of photographs, rather than the subject, many of these biases and selectivity issues are likely avoided. That is, the habitat captured in the iNaturalist photographs is likely to be a 'random sample' of the reef, hence the similarity to habitat captured by the random photo-quadrats, even with the relatively small sample of fish photographs used in this study. However, subject biases may still have some influence on habitat captured incidentally due to potential fish–habitat associations. For example, if parrotfish are photographed more often than less colourful fish, the habitats they associate with may also be over-represented. Such subject biases could potentially have contributed to some habitats being over- or under-represented in this study compared to the standardized photo-quadrats, and this is an area that should be explored further before this technique is broadly implemented.

We suggest that extracting incidental data could be an important ecological monitoring tool, particularly for taxa that are rarely the subject of citizen science photographs. In iNaturalist, for example, as of January 2022 there are 292 000 algae, 39 000 hard coral, 37 000 sponge and 25 000 ascidian photographs globally, in contrast to over 854 000 fish photographs. Here we demonstrated that many of these less-targeted taxa are regularly captured incidentally in the background of popular photographic subjects, such as fish, substantially increasing the observation data available for less charismatic species. While our study highlights the potential of using 'background' data for monitoring marine habitats, our findings could also be applied to many other ecosystems. For example, a similar application of 'background data' was used to investigate plant–pollinator associations by assessing the flowers captured in insect photographs (Bahlai & Landis 2016). Some further extensions of 'incidental' data from citizen science photographs include assessing bird plumage colour (Laitly et al. 2021) and damselfly wing phenotypes (Drury et al. 2019).

The results of this study, while promising, were based on a small selection of sites and a relatively limited number of fish photographs. The slight deviation in the habitat composition captured in opportunistic photographs and standardised surveys at Carrickalinga, for example, could be due to the low number of photographs used in this study. To confirm the validity of using incidental habitat data to monitor reefs, this work should be expanded to include more photographs, sites and times. Ultimately, machine learning should be used to analyse large numbers of photographs to monitor for change in composition or reef health through time at broad spatial scales, with considerable work already underway on automated classification of marine benthic habitats from standardised surveys (e.g. Raphael et al. 2020). In a rapidly changing world, these growing databases, powered by citizen science and machine learning, represent highly promising new tools that can greatly advance environmental monitoring. Here we demonstrated just one of the many potential uses of opportunistic databases such as iNaturalist, with other potential uses including assessing biodiversity, detecting invasive species and determining habitat associations. Ultimately, combining data from opportunistic observations with standardised monitoring (Snäll et al. 2011, Kamp et al. 2016), or with satellite mapping (Leung & Newsam 2010), could allow reef habitats to be monitored rapidly and accurately at broader spatial scales and more frequently than is currently achievable.

*Acknowledgements*. We thank the Australasian Fishes community for their ongoing contributions to iNaturalist and Mark McGrouther and Amanda Hay (Australian Museum) for their work managing the project and for providing the unobscured dataset used for this research. We also thank the Reef Life Survey team and their volunteer divers for collecting the photo-quadrats used in this research. This research was supported by an Australian Government Research Training Program (RTP) Scholarship to C.J.R. and by grant SWR/10/2020 provided by Sea World Research & Rescue Foundation Inc (SWRRFI) and the Winifred Violet Scott Charitable Trust to A.V., C.T.C., A.G.B.P. and C.J.R.; C.T.C. was further supported by a Marie Skłodowska-Curie Individual Fellowship (no. 891052).

## LITERATURE CITED

Aceves-Bueno E, Adeleye AS, Feraud M, Huang Y, Tao M, Yang Y, Anderson SE (2017) The accuracy of citizen science data: a quantitative review. Bull Ecol Soc Am 98: 278–290

Agarwal M (2017) First record of *Dendronotus orientalis* (Baba, 1932) (Nudibranchia: Dendronotidae) in the temperate Eastern Pacific. BioInvasions Rec 6:135–138

Bahlai CA, Landis DA (2016) Predicting plant attractiveness to pollinators with passive crowdsourcing. R Soc Open Sci 3:150677

Callaghan CT, Ozeroff I, Hitchcock C, Chandler M (2020) Capitalizing on opportunistic citizen science data to monitor urban biodiversity: a multi-taxa framework. Biol Conserv 251:108753

Callaghan CT, Poore AGB, Mesaglio T, Moles AT, Nakagawa S (2021) Three frontiers for the future of biodiversity research using citizen science data. Bioscience 71:55–63

Drury JP, Barnes M, Finneran AE, Harris M, Grether GF (2019) Continent-scale phenotype mapping using citizen scientists' photographs. Ecography 42:1436–1445

Edgar GJ, Stuart-Smith RD (2014) Systematic global assessment of reef fish communities by the Reef Life Survey program. Sci Data 1:140007

Kamp J, Oppel S, Heldbjerg H, Nyegaard T, Donald PF (2016) Unstructured citizen science data fail to detect long-term population declines of common birds in Denmark. Divers Distrib 22:1024–1035

Laitly A, Callaghan CT, Delhey K, Cornwell WK (2021) Is color data from citizen science photographs reliable for biodiversity research? Ecol Evol 11:4071–4083

Lenth RV (2021) Emmeans: estimated marginal means, aka least-squares means. R package version 1.7.2. https://CRAN.R-project.org/package=emmeans

Leung D, Newsam S (2010) Proximate sensing: inferring what-is-where from georeferenced photo collections. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Institute of Electrical and Electronic Engineers, San Francisco, CA, p 2955–2962

McKinley DC, Miller-Rushing AJ, Ballard HL, Bonney R and others (2017) Citizen science can improve conservation science, natural resource management, and environmental protection. Biol Conserv 208:15–28

Pocock MJO, Tweddle JC, Savage J, Robinson LD, Roy HE (2017) The diversity and evolution of ecological and environmental citizen science. PLOS ONE 12:e0172579

Poisson AC, McCullough IM, Cheruvelil KS, Elliott KC, Latimore JA, Soranno PA (2020) Quantifying the contribution of citizen science to broad-scale ecological databases. Front Ecol Environ 18:19–26

Prudic KL, Oliver JC, Brown BV, Long EC (2018) Comparisons of citizen science data-gathering approaches to evaluate urban butterfly diversity. Insects 9:186

Raphael A, Dubinsky Z, Iluz D, Netanyahu N (2020) Neural network recognition of marine benthos and corals. Diversity 12:29

Richart CH, Chichester LF, Boyer B, Pearce TA (2019) Rediscovery of the southern California endemic American keeled slug *Anadenulus cockerelli* (Hemphill, 1890) after a 68-year hiatus. J Nat Hist 53:1515–1531

Snäll T, Kindvall O, Nilsson J, Pärt T (2011) Evaluating citizen-based presence data for bird monitoring. Biol Conserv 144:804–810

Thornhill I, Loiselle S, Lind K, Ophof D (2016) The citizen science opportunity for researchers and agencies. BioScience 66:720–721

Troudet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F (2017) Taxonomic bias in biodiversity data and societal preferences. Sci Rep 7:9132

Trygonis V, Sini M (2012) PhotoQuad: a dedicated seabed image processing software, and a comparative error analysis of four photoquadrat methods. J Exp Mar Biol Ecol 424–425:99–108

Uyeda KA, Stow DA, Richart CH (2020) Assessment of volunteered geographic information for vegetation mapping. Environ Monit Assess 192:554

*Editorial responsibility: Thomas Wernberg, Crawley, Western Australia, Australia*
*Reviewed by: D. A. Smerton and 2 anonymous referees*

*Submitted: September 22, 2021*
*Accepted: February 28, 2022*
*Proofs received from author(s): April 13, 2022*