Ecology and Evolution

WILEY

Ecology and Evolution Open Access

RESEARCH ARTICLE OPEN ACCESS

SAMPLE: An R Package to Estimate Sampling Effort for Species' Occurrence Rates

Henrique Bravo[1] | Yacine Ben Chehida[1,2,3] | Sancia E. T. van der Meij[1,4]

[1]GELIFES, University of Groningen, Groningen, the Netherlands | [2]Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Sheffield, UK | [3]Department of Biology, University of York, York, UK | [4]Marine Biodiversity Group, Naturalis Biodiversity Center, Leiden, the Netherlands

Correspondence: Yacine Ben Chehida (y.benchehida@sheffield.ac.uk)

Received: 11 June 2024 | **Revised:** 22 December 2024 | **Accepted:** 4 February 2025

Funding: Funding for fieldwork in Curaçao was financed by the Academy Ecology Fund Grant from the Royal Netherlands Academy of Arts and Sciences (KNAWWF/705/ECO202223), TREUB-maatschappij (Society for the Advancement of Research in the Tropics), FONA Foundation for Research on Nature Conservation and Stichting Fonds Dr. Christine Buisman. Yacine Ben Chehida was supported by a grant from the Natural Environment Research Council (NERC, NE/T008121/1) from the University of Sheffield and the University of York.

Keywords: community ecology | prevalence | sample size | simulations | symbiont

ABSTRACT

Species' occurrence rates are the backbone of many ecological studies. Sampling of species occurrence, however, can come with challenges and might prove more difficult than anticipated. Logistical difficulties, limited funds or time, elusiveness or rarity of species and difficult sampling environments are all examples of scenarios that might contribute to (undesired) small sample sizes. In order to help circumvent some of these difficulties and uncertainties, we present SAMPLE, an R package that aims to inform the user whether the amount of sampling conducted up until a chosen moment is enough to accurately estimate the occurrence rate of species. We use a simulation approach to help verify the accuracy of the package and to help guide the user in choosing the most appropriate values for the available parameters. Moreover, we provide a real data set where we used SAMPLE to estimate the occurrence rate of various coral-dwelling species on their hosts and the minimum number of samples required for an accurate estimation. This provided example data set includes closely related host species, single or multiple symbionts on a single host coral taxon, and data points obtained from different depths to illustrate how occurrence rates can vary depending on the provided input. Due to its simplicity and ease of use, this package allows users to run it while in the field to estimate if sampling is sufficient or if the sampling approach needs to be adapted for a particular species. We hope that this package proves itself useful to users that need to estimate occurrence or prevalence rates of species and do not always have the possibility to obtain large sample sizes.

1 | Introduction

Sample size is an important factor in any ecological research project. Large sample sizes improve the precision of estimations and the power of statistical tests, but also increase the costs and length of a field campaign (Underwood 1997; Singh and Masuku 2014). Study design typically includes estimations of sampling effort based on experience (Kenkel et al. 1989; Schreiber and Brauns 2010), experimental design (Bernstein and Zalinski 1983; Underwood and Chapman 2003) and possibly, data simulations (Nadon and Stirling 2006; Guerra-Castro et al. 2021). To study biodiversity patterns, one of the most frequent types of ecological data collected is the occurrence of species (i.e., presence/absence) and how it relates to environmental and geographical patterns (Soberón and Peterson 2004). This data can be useful, for example, towards studies on

The first two authors contributed equally to this article.

habitat suitability of species (Hirzel et al. 2006), biogeography (Hanski 1982), conservation (Rondinini et al. 2006) or community ecology (Gotelli 2004). Prevalence (rate) is a measure of occurrence (commonly used in parasitology and disease ecology) and is used here as the mean proportion of host taxa inhabited by a symbiont (Bush et al. 1997).

A large sample size is usually preferable when it comes to accurately determining the occurrence or prevalence rate of species (Gregory and Blackburn 1991; Gotelli and Colwell 2001). The relationship between sample size and accurate estimation of rates is not, however, a linear one, and it might not be necessary to carry out an exhaustive sampling effort to obtain a sample size that is representative of the studied community (Gregory and Woolhouse 1993; Chao and Bunge 2002). The uncertainty of occurrence and prevalence rates of species rapidly decreases after a minimum number of samples has been reached (e.g., 10–20 individuals) and stabilises shortly after (Underwood 1997; Jovani and Tella 2006). In situations where it can be difficult to increase the sampling effort due to, for example, limited time or funds, logistical difficulties or elusiveness of study species, it is essential to know whether the effort carried out is enough or not.

There are statistical methods that have been developed to cope with reduced or unbalanced sample sizes (Tella et al. 1999; Paterson and Lello 2003; Jovani and Tella 2006). Moreover, specific statistical approaches, R packages, and online calculators have been developed to predict the amount of sampling effort needed in a study, but they tend to require the provision of input data that will help in that sampling effort's estimation (Anderson and Santana-Garcon 2015; Guerra-Castro et al. 2021; Manitz et al. 2020; Naing et al. 2022). The required information can consist of, for example, the known or expected occurrence/prevalence rate (Manitz et al. 2020; Naing et al. 2022) or a pilot data set to help guide the user in choosing an appropriate sampling effort (Anderson and Santana-Garcon 2015; Guerra-Castro et al. 2021).

Given the usefulness, but also the difficulty, of knowing what the necessary sampling effort is for having accurate occurrence/prevalence rates of species without providing any initial input, we developed an R package to help with such estimations: SAMPLE. This package allows the user to determine whether the amount of sampling conducted so far is sufficient, or whether specific taxa are underrepresented in the data set. It has the added value that due to its simplicity and ease of use, it can be run while in the field. All of the parameters used in this package have set default values that should be the most useful for the majority of users, but they can be manually adjusted to fit the different needs of users and their study environments.

This package can be used for all kinds of terrestrial and marine community assemblages; however, as an example, we simulated data on hypothetical host organisms and their associated symbiont species to estimate the prevalence rates of the symbionts on their hosts in order to verify the accuracy in prevalence rate estimation of the SAMPLE R package. This simulation takes known prevalence rates of symbionts in populations of different sizes and tries to estimate the prevalence rates of these populations using different combinations of parameters (e.g., replicate number, different number of successive points).

Additionally, this paper uses a real data set of different species of obligate coral-dwelling faunal symbionts of stony corals and hydrocorals (see Figure 1 for an example). This data set allows to further test if the prevalence rate of symbionts across different coral host species was accurately estimated, and if so, what would have been the minimum sampling effort needed.

2 | Functionality of SAMPLE

The SAMPLE R package presented here takes as input a data set that can be accessed directly in the package or in Data S1, where each row corresponds to a single data point. In the provided example, data prevalence rates are calculated with the names of species in the first column corresponding to the hosts, and their symbiont(s)' occurrence(s) noted down with 1 (=presence) and 0 (=absence) in the remainder of the columns (i.e., each column representing a different symbiont species). To estimate, for example, the occurrence rates of species across habitats, the first column should contain the different habitats instead of host species. The rows should denote the occurrence of the species of choice across different habitats with 1 (i.e., presence) and 0 (i.e., absence) as values.

The data set is subsequently exposed to a process of random sampling without replacement, where a sample of k individuals from the initial data set is randomly sampled and its prevalence rate calculated with k ranging between 1 and k_{MAX} (i.e., the maximum number of available individuals). The process is repeated n times (user defined, default $n = 50$) in order to obtain an average occurrence or prevalence rate with k samples over n iterations.

The stability point of the occurrence or prevalence rate (i.e., the minimum number of samples needed to accurately determine the occurrence or prevalence rate of the study system) is determined by first computing the difference of x successive (mean) prevalence rates (user defined, default = 10) that are below a threshold y (user defined, default $y = 2$). Because y values become smaller as k increases, y is divided by the square root of k (i.e. $\frac{y}{\sqrt{k}}$) to correct for the effect of sample size. The difference between the minimum and maximum values among the x means

FIGURE 1 | The symbiont *Spirobranchus giganteus* (also commonly known as Christmas tree worm) on a colony of the brain coral *Pseudodiploria strigosa*. Photo by S.E.T. van der Meij.

(Delta: Δ) is then computed. The stability point is set as the first of the x successive means below a z threshold (user defined, default $z = 1$).

A schematic representation of this explanation can be found in Figure 2. This is then repeated for each host and each symbiont species, with a plot being generated at the end (see section on Visualisation and Figure 3 for an example).

3 | Simulated Data: Prevalence Rates in a Host/Symbiont System

We simulated a host/symbiont data set with known population sizes and prevalence rates in order to test the validity and accuracy of the SAMPLE R package. Three population sizes were chosen (i.e., 100, 1000 and 10,000) and prevalence rates of 10%–90% with increments of 10% were then calculated for these different populations. For each combination of prevalence rate and population size, different numbers of replicates were used, ranging from 10 to 100 (also in increments of 10), then 200 and 500 (see Figure S1A for a schematic representation of the different simulations that were conducted). These simulations were all run with the default values (i.e., successive points = 10; mean-difference = 2; Δ = 1). An example of a population with 1000 individuals, a prevalence rate of 50%, and 50 replicates was then taken and further parameters were tested. The number of successive points (i.e., 2, 10 and 50), mean-difference (i.e., 1, 2, 5 and 10) and Δ values (i.e., 0.5, 1 and 2) were all tested in combination with each other (Figure S1B). One final example was taken with set parameters and settings and was run 10 times in order to evaluate the natural stochasticity of the process.

The first part of the simulations dealt with different population sizes and prevalence rates, as well as varying numbers of replicates. Default values for the other parameters revealed that, with few exceptions, all the estimated prevalence rates were within 1% of the actual prevalence rates (Table S1). The number of samples required to estimate the prevalence rate in all of our simulations was, on average, 32, but this value was as low as 20 for prevalence rates of 10% or 90% and as high as 40 samples for prevalence rates of 40%–60%. This natural variation is a

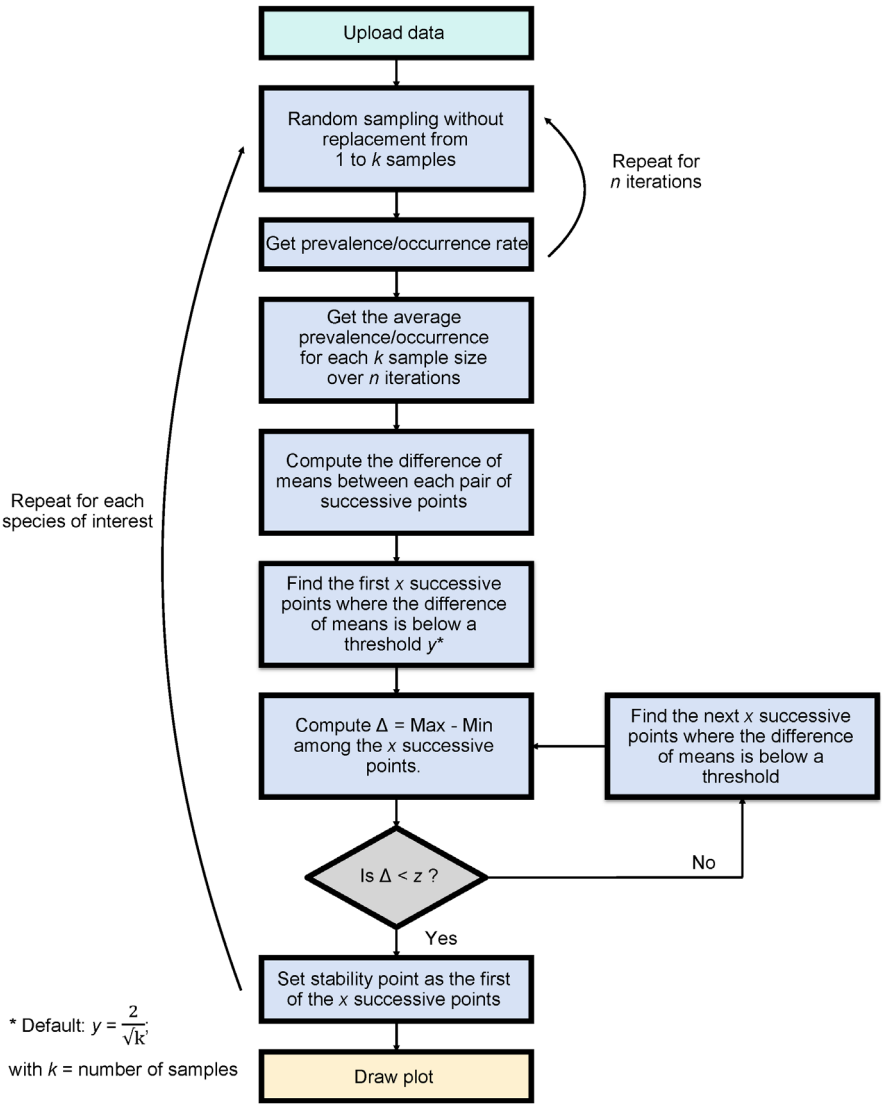

FIGURE 2 | Schematic representation of the functionality of the SAMPLE R package.

direct consequence of the variance of the binomial-like nature of the sampling process (see Texts S2–S3 for further information). Comparable results were obtained using 1000 replicates for both the number of samples necessary to reach stability (Figure S2) and the error between the prevalence estimated by SAMPLE and the known simulated prevalence (Figure S3).

The second part of the simulations, where we changed the number of successive points, the value of mean-difference and the Δ value revealed more substantial variations in the results, highlighting the importance of choosing meaningful values. A higher number of successive points naturally resulted in a higher number of samples needed, as did a smaller value (i.e., 0.5) for Δ and smaller values of mean-difference (e.g., the lower the value the more samples required).

We then ran a final set of simulations 10 times using the default values for all parameters (i.e., successive points = 10; mean-difference = 2; Δ = 1; number of replicates = 50) on a population with 1000 individuals and a prevalence rate of 50%. Even though there was only one instance in which the predicted prevalence rate varied from the actual one by more than 1%, the number of samples required to estimate it was on average 38 ± 9.5, highlighting the natural variation of the process. We therefore

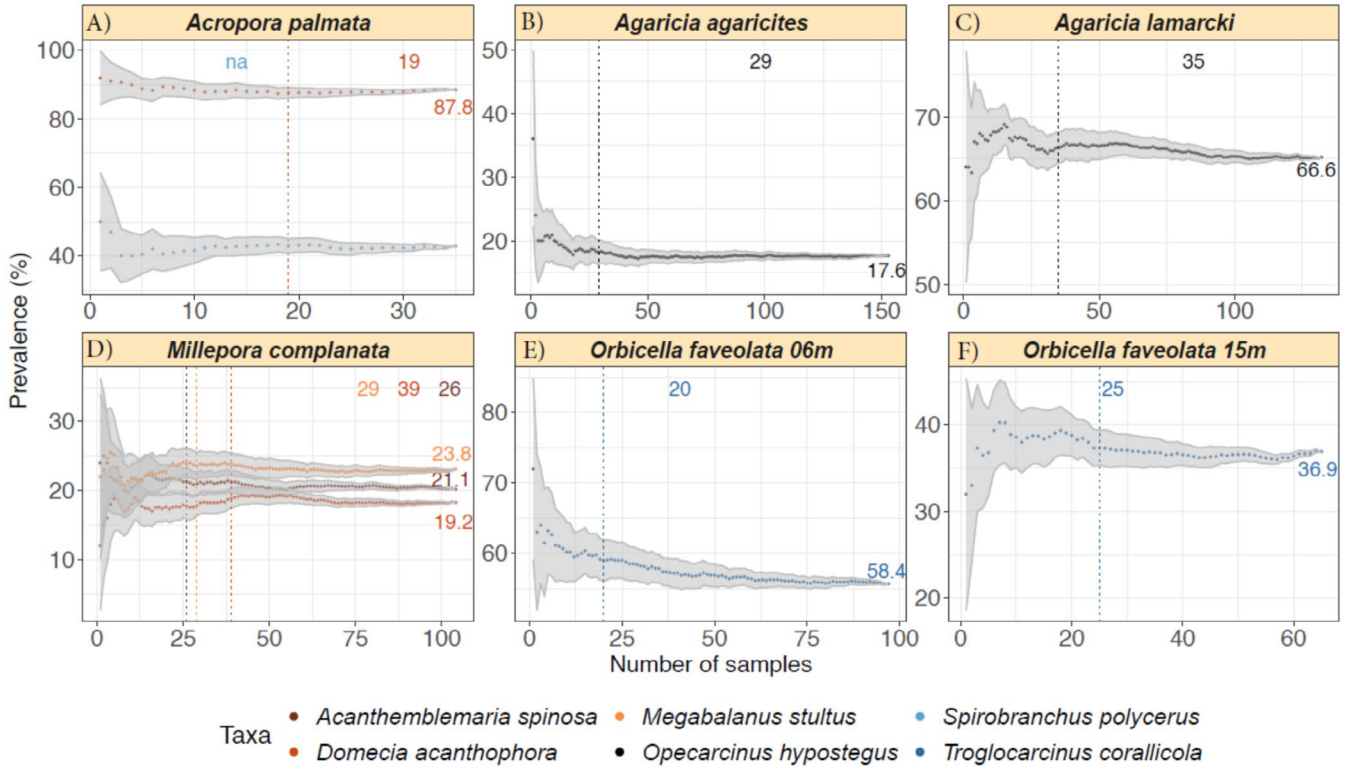

FIGURE 3 | Output plot of SAMPLE depicting the minimum number of samples needed for the estimation of stable prevalence rates of different species of symbionts on their respective coral hosts. For a more detailed explanation of the different elements on the plot, please refer to Figure 4.

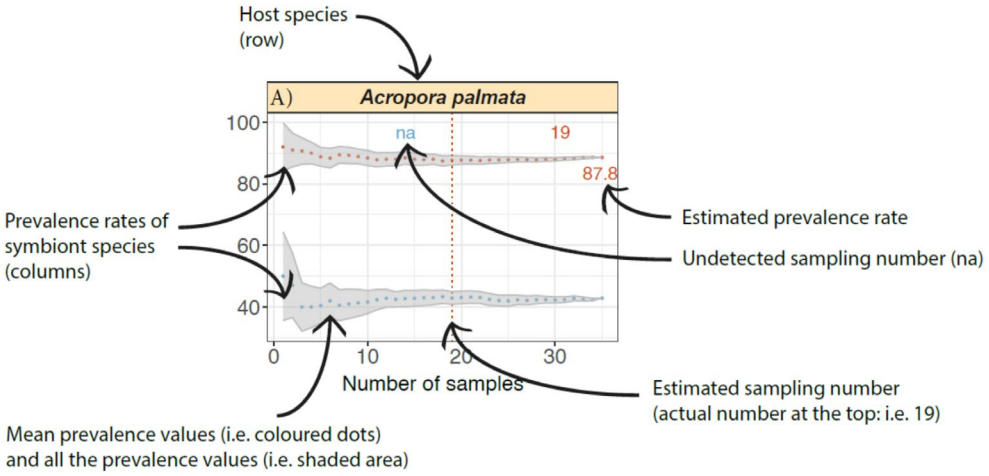

FIGURE 4 | A detailed explanation of what the different elements present on the output plot represent.

suggest users to run the analysis on their real data sets a minimum of five times to avoid getting a number of required samples that by chance deviates far from the mean.

4 | Real Data: Prevalence Rates of Coral-Dwelling Fauna

In order to test the accuracy of the R package, we ran SAMPLE on a real data set of corals and their associated dwelling fauna, sampled from 586 host colonies at various sites along the leeward side of Curaçao (Dutch Caribbean) between the 24th February and the 30th March 2022.

Four species of stony corals (Scleractinia) and one hydrocoral (Hydrozoa) species were selected for the prediction of sampling effort needed for the accurate estimation of symbiont prevalence rate: (A) the stony coral *Acropora palmata* with the crab *Domecia acanthophora* and fanworm *Spirobranchus polycerus*; (B) the stony coral *Agaricia agaricites* with the crab *Opecarcinus hypostegus*; (C) the stony coral *Agaricia lamarcki* with *O. hypostegus*; (D) the hydrocoral *Millepora complanata* with the barnacle *Megabalanus stultus*, *D. acanthophora* and the blenny *Acanthemblemaria spinosa*; (E) the stony coral *Orbicella faveolata* with the crab *Troglocarcinus corallicola* at 6 m; and (F) *Orbicella faveolata* with *T. corallicola* at 15 m. This data set will enable us to detect patterns of (1) prevalence of multiple symbiont species on a single host (A and D); (2) prevalence of the same symbiont on closely related hosts (B and C); and (3) prevalence of the same symbiont/host pair across two different depths (E and F).

Using the aforementioned coral and associated fauna data set, included in the SAMPLE package and in Data S1, we were able to estimate the minimum number of sampled corals needed to accurately estimate the prevalence rate of their symbionts. The default visualisation output of the package is a panel figure where the number of panels is equal to the number of host species from the input data set. The title of each panel corresponds to the name of the host species, and the names of the symbionts are provided in the legend at the bottom of the figure (Figure 3). The estimated prevalence rate is provided on the right-hand side of each panel, next to the tail end of the prevalence curve, and the number of samples needed to estimate that same prevalence rate is shown at the top of the panel and indicated with a dotted line, both in the same colour as the symbiont in question (see Figure 4 for a more detailed explanation). For example, the prevalence rates of the blenny *Acanthemblemaria spinosa*, the crab *Domecia acanthophora,* and the barnacle *Megabalanus stultus* on the hydrocoral *Millepora complanata* are 21.1%, 19.2%, and 23.8%, respectively. The number of sampled *M. complanata* hydrocorals needed to obtain those prevalence rates was 26, 39, and 29 (Figure 3D), respectively, highlighting that our sampling effort ($n = 104$) was sufficient. In the case of multiple symbionts requiring the exact same number of samples, only one dotted line will be presented, but both numbers will be printed. If the number of samples is not enough to estimate the prevalence rate, then *na* is printed at the top of the graph in the colour of the respective symbiont, as can be seen for *Spirobranchus polycerus* on *Acropora palmata*. In this case, SAMPLE indicates that our sampling effort ($n = 35$) was too low to obtain a meaningful prevalence rate for that particular symbiont (Figure 3A). However, Figure 3A shows that stabilisation is close, but likely has not yet been reached because the number of successive points required surpasses the number of data points obtained.

5 | Discussion

Reliable estimations of occurrence or prevalence rates are not always easy to achieve. It has been suggested that the more samples collected, the more reliable the estimations (Gregory and Blackburn 1991), and while big sample sizes tend to be preferred, this is not always necessary or possible (Underwood 1997; Jovani and Tella 2006). There are many factors that can contribute to small sample sizes: the sampling environment can be difficult to access (e.g., deep sea or cave environments, political instability in certain countries), sampling time can be limited (e.g., funding, scuba diving, boat time), or specimens are elusive and/or rare. Knowing whether the samples collected allow for a representative estimation of occurrence rates is therefore needed when sampling time and/or funds are limited.

The SAMPLE R package provided here aims at reducing this uncertainty and informing the user of the minimum number of samples needed to accurately estimate occurrence/prevalence rates, or otherwise informing the user that more sampling is required. SAMPLE can also be run during field campaigns, highlighting which taxa ideally need more data points. Even in situations where the sampling conducted is not enough to estimate a prevalence rate, and more data cannot be obtained, the visual output (see Figure 3A as an example) allows the user to make a prediction as to whether a stabilisation of the prevalence rate is close or not.

This package can detect an early stabilisation of the prevalence rate of symbionts, and this is the value that is reported initially. It is possible, albeit unlikely, that another prevalence rate will be found in that same host/symbiont system that differs from the initial one if more runs are done; this is due to the stochastic nature of the process (see Figures S5–S6 and Texts S1–S3). Parameters can, however, be manually changed in order to make the estimation of prevalence rate and number of samples needed more or less stringent. Moreover, small variations in prevalence rate values (i.e., less than 0.1%) are not necessarily ecologically relevant (Jovani and Tella 2006), and the variation found in the simulation examples provided here tends to vary little (i.e., less than 1% in most cases). Ultimately it is up to the user to determine the value that is the most ecologically relevant for their particular study system and to determine the minimum sample size required to estimate it. We recommend having a careful look at the results of the simulation (see Table S1 and Text S1) to understand how changing certain parameters may affect the estimated prevalence rates and the number of samples required and then, if necessary, adjust the parameters to fit specific needs.

One of the strengths of this R package is also one of its weaknesses: being able to set so many parameters allows users to choose the properties that suit them best, leading to robust results. On the other hand, it also means that if users do not know their study species or system well, they might end up choosing

parameters that are not the most suitable. The default values should, however, be suitable for most study systems. The nature of this bootstrapping approach also comes with some intrinsic variation, so there will be some stochasticity in the results (Figures S2–S4 and Texts S2–S3). As mentioned before, a way of counteracting this is by repeatedly running the analysis (we suggest a minimum of five runs) so the results obtained can converge, but there will always be some variation (Figures S2–S4 and Text S1). Lastly, some prevalence rates can be harder to estimate, particularly those that are very high or very low (Gregory and Blackburn 1991; Jovani and Tella 2006), as highlighted by the results of our simulation, so when possible, users should be less conservative with their sampling effort to account for this.

It is important to note that prevalence rates of symbionts across their hosts, even when looking at the same host/symbiont system, can change across time and space (Shykoff and Kaltz 1998; Penczykowski et al. 2016; Starkloff and Galen 2023). In the example provided above, we looked at the same species of symbiont, the crab *T. corallicola*, on the same coral host, *O. faveolata*, across two different depths (6 and 15 m). SAMPLE detected different prevalence rates and slightly different numbers of samples needed ($n = 20$ and 25, respectively) for the estimation of prevalence rates (Figure 3E,F). We also estimated the prevalence rates of the same symbiont, the crab *O. hypostegus*, on two closely related coral host species, *A. agaricites* and *A. lamarcki*, across the same depth range (i.e., from 6 to 15 m) and the results varied considerably (Figure 3B,C). These examples highlight that ecological factors impact symbiont prevalence rates across different depths and hosts (van Tienderen and van der Meij 2016). Ultimately, it is up to the user to choose when and how to use this package to best suit their needs (Text S1 and S3).

Finally, it should also be mentioned that this package is not aimed at determining overall prevalence rates for host/symbiont systems, or overall occurrence rates of species. It is rather aimed at determining whether the sampling effort conducted is enough to estimate a stable prevalence rate for that particular group of taxa at that particular point in time.

Author Contributions

Henrique Bravo: conceptualization (equal), data curation (lead), formal analysis (lead), funding acquisition (equal), investigation (equal), methodology (equal), project administration, resources, software (supporting), supervision (equal), validation, visualization, writing – original draft (lead), writing – review and editing. **Yacine Ben Chehida:** conceptualization (equal), formal analysis (equal), investigation (supporting), methodology (equal), software (lead), validation (equal), visualization (equal), writing – original draft (supporting). **Sancia E. T. van der Meij:** conceptualization (supporting), data curation (supporting), funding acquisition (lead), investigation (supporting), project administration (equal), resources (equal), supervision (lead), validation (equal), writing – review and editing (equal).

Acknowledgements

The data collection in the field was a joint effort conducted by the first author, Natascha Borgstein, Tao Xu and Yun Scholten (all University of Groningen). Without their help, there would be no real data set for this study. We thank Pedro Santos Neves and Raphaël Scherrer, also from the University of Groningen, for their help in compiling the original code into the SAMPLE R package. We further thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Hábrók high-performance computing cluster. A final thank you to the CARMABI research station on Curaçao for helping us with logistical support.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The version of the SAMPLE R package used in this study is archived and available from the online Zenodo repository (version 1.0.0; https://doi.org/10.5281/zenodo.13869566). Regularly updated versions of the package and help files are available from CRAN and on GitHub (https://github.com/yacinebenchehida/SAMPLE). The R code and explanations from the simulation work that contributed to the development of this R package are provided as online Supplementary material in this paper, and in the aforementioned GitHub repository.

References

Anderson, M. J., and J. Santana-Garcon. 2015. "Measures of Precision for Dissimilarity-Based Multivariate Analysis of Ecological Communities." *Ecology Letters* 18, no. 1: 66–73.

Bernstein, B. B., and J. Zalinski. 1983. "An Optimum Sampling Design and Power Tests for Environmental Biologists." *Journal of Environmental Management* 16: 35–43.

Bush, A. O., K. D. Lafferty, J. M. Lotz, and A. W. Shostak. 1997. "Parasitology Meets Ecology on Its Own Terms: Margolis Et al. Revisited." *Journal of Parasitology* 83, no. 4: 575–583.

Chao, A., and J. Bunge. 2002. "Estimating the Number of Species in a Stochastic Abundance Model." *Biometrics* 58: 531–539.

Gotelli, N. J. 2004. "A Taxonomic Wish-List for Community Ecology." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 359, no. 1444: 585–597.

Gotelli, N. J., and R. K. Colwell. 2001. "Quantifying Biodiversity: Procedures and Pitfalls in the Measurement and Comparison of Species Richness." *Ecology Letters* 4: 379–391.

Gregory, R. D., and T. M. Blackburn. 1991. "Parasite Prevalence and Host Sample Size." *Parasitology Today* 7, no. 11: 316–318.

Gregory, R. D., and M. E. Woolhouse. 1993. "Quantification of Parasite Aggregation: A Simulation Study." *Acta Tropica* 54, no. 2: 131–139.

Guerra-Castro, E. J., J. C. Cajas, N. Simões, J. J. Cruz-Motta, and M. Mascaró. 2021. "SSP: An R Package to Estimate Sampling Effort in Studies of Ecological Communities." *Ecography* 44, no. 4: 561–573.

Hanski, I. 1982. "Dynamics of Regional Distribution: The Core and Satellite Species Hypothesis." *Oikos* 38, no. 2: 210–221.

Hirzel, A. H., G. Le Lay, V. Helfer, C. Randin, and A. Guisan. 2006. "Evaluating the Ability of Habitat Suitability Models to Predict Species Presences." *Ecological Modelling* 199, no. 2: 142–152.

Jovani, R., and J. L. Tella. 2006. "Parasite Prevalence and Sample Size: Misconceptions and Solutions." *Trends in Parasitology* 22, no. 5: 214–218.

Kenkel, N. C., P. Juhász-Nagy, and J. Podani. 1989. "On Sampling Procedures in Population and Community Ecology." *Vegetatio* 83, no. 1–2: 195–207.

Manitz, J., M. Hempelmann, G. Kauermann, et al. 2020. *samplingbook: Survey Sampling Procedures*. R package version 1.2.4. https://CRAN.R-project.org/package=samplingbook.

Nadon, M. O., and G. Stirling. 2006. "Field and Simulation Analyses of Visual Methods for Sampling Coral Cover." *Coral Reefs* 25, no. 2: 177–185.

Naing, L., R. B. Nordin, H. Abdul Rahman, and Y. T. Naing. 2022. "Sample Size Calculation for Prevalence Studies Using Scalex and ScalaR Calculators." *BMC Medical Research Methodology* 22, no. 209: 1–8. https://doi.org/10.1186/s12874-022-01694-7.

Paterson, S., and J. Lello. 2003. "Mixed Models: Getting the Best Use of Parasitological Data." *Trends in Parasitology* 19, no. 8: 370–375.

Penczykowski, R. M., A. L. Laine, and B. Koskella. 2016. "Understanding the Ecology and Evolution of Host–Parasite Interactions Across Scales." *Evolutionary Applications* 9, no. 1: 37–52.

Rondinini, C., K. A. Wilson, L. Boitani, H. Grantham, and H. P. Possingham. 2006. "Tradeoffs of Different Types of Species Occurrence Data for Use in Systematic Conservation Planning." *Ecology Letters* 9, no. 10: 1136–1145.

Schreiber, J., and M. Brauns. 2010. "How Much Is Enough? Adequate Sample Size for Littoral Macroinvertebrates in Lowland Lakes." *Hydrobiologia* 649, no. 1: 365–373.

Shykoff, J., and O. Kaltz. 1998. "Local Adaptation in Host–Parasite Systems." *Heredity* 81: 361–370.

Singh, A. S., and M. B. Masuku. 2014. "Sampling Techniques & Determination of Sample Size in Applied Statistics Research." *International Journal of Economics, Commerce and Management* 2, no. 11: 1–22.

Soberón, J., and A. T. Peterson. 2004. "Biodiversity Informatics: Managing and Applying Primary Biodiversity Data." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 359, no. 1444: 689–698.

Starkloff, N. C., and S. C. Galen. 2023. "Coinfection Rates of Avian Blood Parasites Increase With Latitude in Parapatric Host Species." *Parasitology* 150, no. 4: 329–336.

Tella, J. L., G. Blanco, M. G. Forero, A. Gajón, J. A. Donázar, and F. Hiraldo. 1999. "Habitat, World Geographic Range, and Embryonic Development of Hosts Explain the Prevalence of Avian Hematozoa at Small Spatial and Phylogenetic Scales." *Proceedings of the National Academy of Sciences of the United States of America* 96, no. 4: 1785–1789.

van Tienderen, K. M., and S. E. T. van der Meij. 2016. "Occurrence Patterns of Coral-Dwelling Gall Crabs (Cryptochiridae) Over Depth Intervals in the Caribbean." *PeerJ* 4: e1794.

Underwood, A. J. 1997. *Experiments in Ecology: Their Logical Design and Interpretation Using Analysis of Variance*. Cambridge University Press.

Underwood, A. J., and M. G. Chapman. 2003. "Power, Precaution, Type II Error and Sampling Design in Assessment of Environmental Impacts." *Journal of Experimental Marine Biology and Ecology* 296, no. 1: 49–70.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.