

Methods in Ecology and Evolution

DR JUAN ANTONIO BALBUENA (Orcid ID : 0000-0003-4006-1353)

MISS CRISTINA LLOPIS-BELENGUER (Orcid ID : 0000-0001-7672-0088)

DR ISABEL BLASCO-COSTA (Orcid ID : 0000-0002-5001-5986)

Article type : Application

Handling editor: Dr Will Pearse

Fuzzy Quantification of Common and Rare Species in Ecological Communities (FuzzyQ)

Juan A. Balbuena¹, Clara Monlleó-Borrull¹, Cristina Llopis-Belenguer¹, Isabel Blasco-Costa^{2,3}, Volodimir L. Sarabeev⁴, Serge Morand^{5,6}

¹ Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, Valencia, Spain

² Department of Invertebrates, Natural History Museum of Geneva, Geneva, Switzerland

³ Department of Arctic and Marine Biology, UiT The Arctic University of Norway, Tromsø, Norway

⁴ Department of Biology, Zaporizhzhia National University, Zaporizhzhia, Ukraine

⁵ Institut des Sciences de l'Evolution, CNRS, CC065, Université de Montpellier, Montpellier Cedex, France; CIRAD ASTRE

⁶ Faculty of Veterinary Technology, Kasetsart University, 10400 Bangkok, Thailand

Correspondence: Juan A. Balbuena Email: j.a.balbuena@uv.es

Running headline: Fuzzy Quantification of Common and Rare Species

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/2041-210X.13588](https://doi.org/10.1111/2041-210X.13588)

This article is protected by copyright. All rights reserved

Abstract

1. Most species in ecological communities are rare whereas only a few are common. This distributional paradox has intrigued ecologists for decades but the interpretation of species abundance distributions remains elusive.
2. We present Fuzzy Quantification of Common and Rare Species in Ecological Communities (FuzzyQ) as an R package. FuzzyQ shifts the focus from the prevailing species-categorization approach to develop a quantitative framework that seeks to place each species along a rarity-commonness gradient. Given a community surveyed over a number of sites, quadrats, or any other convenient sampling unit, FuzzyQ uses a fuzzy clustering algorithm that estimates a probability for each species to be common or rare based on abundance-occupancy information. Such a probability can be interpreted as a commonness index ranging from 0 to 1. FuzzyQ also provides community-level metrics about the coherence of the allocation of species into the common and rare clusters that are informative of the nature of the community under study.
3. The functionality of FuzzyQ is shown with two real datasets. We demonstrate how FuzzyQ can effectively be used to monitor and model spatio-temporal changes in species commonness, and assess the impact of species introductions on ecological communities. We also show that the approach works satisfactorily with a wide range of communities varying in species richness, dispersion and abundance currencies.
4. FuzzyQ produces ecological indicators easy to measure and interpret that can give both clear, actionable insights into the nature of ecological communities and provides a powerful way to monitor environmental change on ecosystems. Comparison among communities is greatly facilitated by the fact that the method is relatively independent of the number of sites or sampling units considered. Thus, we consider FuzzyQ as a potentially valuable analytical tool in community ecology and conservation biology.

Key-Words: Rarity, commonness, abundance-occupancy distributions, fuzzy clustering, community ecology, assembly rules, environmental monitoring.

1 INTRODUCTION

Ecological communities are formed by species that differ widely in abundance. Almost invariably the observation is that most species are rare, whereas a few are common (Magurran & Henderson, 2011). This pervasive pattern has intrigued ecologists for decades but, despite the large literature on the topic, the interpretation of species abundance distributions remains elusive (Werner et al., 2014; Enquist et al., 2019). The assumption often made is that underlying factors, such as

immigration, succession and competition, eventually determine differences in establishment and persistence of each species in the community (McGill et al., 2007; McGill, 2011; Alroy, 2015; Calatayud et al., 2019).

A quantitative framework for species commonness and rarity amenable to hypothesis testing and statistical modelling would facilitate evaluating the roles played by demographic variables and species traits, thereby illuminating assembly rules in ecological communities. Such a framework would also be extremely valuable for conservation biology in at least three important aspects: (1) Species rarity is an important predictor of extinction risk because the impact of environmental disturbances is expected to be higher on small populations (Davies, Margules, & Lawrence, 2000) and since budgets for biodiversity conservation are limited, quantifying the rarity of species would facilitate prioritizing some over the others. (2) From a functional perspective, the role played by common and rare species in providing ecosystem services is currently under scrutiny (Leitão et al., 2016, Loiseau et al., 2020). Although intuitively it might be assumed that most of the ecosystem functionality should reside in the former, the contribution of rare species is still poorly understood and, in fact, might be substantial (Leitão et al., 2016; Violle et al., 2017; Dee et al., 2019, Loiseau et al., 2020). So assessing rarity could also be justified in terms of identifying species that provide essential ecosystem services (Flather & Sieg, 2007; Violle et al., 2017; Dee et al., 2019, Loiseau et al., 2020) or stabilize ecological communities (Calatayud et al., 2019). (3) Monitoring variation of commonness-rarity patterns over time or along geographical and environmental gradients provides a simple way to obtain crucial information on ecosystem changes (McGill, 2011). For instance, if common species become increasingly rare in response to environmental disturbances, it might have a cascading effect on the rest of the community (Gaston & Fuller, 2008).

Thus, metrics of commonness and rarity at species and community level would be extremely useful to unveil the architecture of ecological communities, assess the likelihood of extinction of rare species, correlate commonness or rarity with functional distinctiveness and monitor environmental change. However, a universal quantitative framework is currently lacking. A great deal of effort has been put on establishing the distribution patterns emerging from the categorization of species as common or rare (Gray, Bjørgesæter, & Ugland, 2005; McGill et al., 2007; Antão, Connolly, Magurran, Soares, & Dornelas, 2017). However, a major problem of fitting models to species abundance distributions has been adjusting the data to a suitable theoretical distribution (Williamson & Gaston, 2005; McGill et al., 2007; Alroy, 2015). To some

extent this is because the border between common and rare species is often blurred (Magurran & Henderson, 2011), which has led authors to propose additional subcategories of rarity (Rabinowitz, 1981; Hanski, 1991; Yu & Dobson, 2000; Arnan, Gaucherel, & Andersen, 2011).

Herein we propose shifting the focus from species categorization to a quantitative approach to place each species along a rare-commonness gradient. This will be achieved by fuzzy clustering, which is ideal for handling scenarios in which borderline observations and ambiguity challenge traditional binary classification (Dick & Laflamme, 2018). A distinct feature of fuzzy classification is that each observation is assigned to each cluster considered with a different level of certainty, which is measured by membership coefficients (Kaufman & Rousseeuw, 1990). This quantification of ambiguity facilitates the analysis of complex patterns in natural systems and not surprisingly fuzzy classification approaches are being increasingly applied in ecology (e.g. Barbosa, 2015; Fiorentino et al., 2017; Dick & Laflamme, 2018; Bagnaro et al., 2020).

Fuzzy Quantification of Common and Rare Species in Ecological Communities (FuzzyQ) is based on the analysis of the abundance-occupancy (AO) relationship of species in a community, which assumes a positive relationship between local abundance and occupancy (Gaston et al., 2000; Gaston & He, 2011). Given a community surveyed over a number of sites, quadrats, or any other convenient sampling unit, FuzzyQ applies fuzzy clustering to estimate a probability for each species to be common or rare based on its AO.

Although widely used, we acknowledge at the onset that abundance and/or occupancy are not the only criteria to assess species commonness and rarity (Rabinowitz, 1981; Gaston, 1994, 1997). However, the key point is that regardless of the data used, we can always use fuzzy clustering to quantify the degree of belonging of each species to the common or rare categories (or any other pre-established categorization for that matter).

We show herein that FuzzyQ produces ecological indicators easy to measure and interpret that are amenable to hypothesis testing and statistical modelling. In addition, FuzzyQ is distribution free, i.e. no a priori assumption about the distribution of species abundances is required. We illustrate the capabilities of the framework with two real-world examples involving each related and unrelated (i.e. not sharing species) communities and evaluate the effect of sample size on the estimation of commonness and rarity.

2 OVERVIEW OF FUZZYQ

FuzzyQ evaluates simultaneously the dissimilarities in occupancy and abundance among species in a community and applies fuzzy clustering to allocate them into two clusters of rare and common

species. Since occupancy and abundance are measured in different scales and can come in different units (for instance, the former can be reported as either number or fraction of sites occupied), FuzzyQ assesses by default Gower's (1971) dissimilarities, which are appropriate for such mixed data. In this framework, the cluster membership coefficients produced are re-interpreted as indices of commonness (C_i) and rarity (R_i) that reflect the probability of species i being common and rare, respectively. In addition, FuzzyQ returns global metrics that inform of the coherence of each cluster and the strength of the separation between common and rare species in the community.

FuzzyQ is provided as an R package (R Core Team, 2020), available on CRAN (<https://CRAN.R-project.org/package=FuzzyQ>). (See Supplementary Information for details about the fuzzy clustering algorithms and FuzzyQ functions). We first illustrate the application of FuzzyQ with a dataset of ant species (`ants_Darwin_A` in Calatayud et al., 2019) that collates the abundance of 46 species in 100, 18×18 m plots sampled in the Northern Territory, Australia (Arnan et al., 2011). Below we provide an outline of the main steps of the analysis. The full R session is available at https://ligophorus.github.io/FuzzyQ/ants_example.html.

We first load the library and the dataset:

```
> library(FuzzyQ)
> data(antsA)
```

where `antsA` is a site-by-species abundance matrix.

Function `fuzzyq` applies fuzzy clustering (Eq. 1 Supplementary Information) to `antsA`:

```
> FQAnts <- fuzzyq(antsA, sorting = TRUE)
```

For effective visualization of the rare and common species composition of the community, we set `sorting = TRUE`, which sorts species results by silhouette width (see below) within each cluster (Fig. 1). `FQAnts` includes three main pieces of information:

(1) `FQAnts$AO` lists the AO data per species (given as relative frequency of sites occupied and mean abundance per site) (Fig. 1a).

(2) `FQAnts$spp` contains the fuzzy clustering results per species. The R-base function `head` returns those of the first six species:

```
> head(FQAnts$spp)
      cluster sil_width Common.I
sps_15      0 0.4931739 0.4519339
sps_8       0 0.5127858 0.4370737
sps_45      0 0.6232735 0.3602032
```

```
sps_35      0 0.6520166 0.3366034
sps_18      0 0.8040307 0.1797200
sps_19      0 0.8336707 0.1426252
```

`cluster` is a binary variable that codes the allocation of species to the rare (0) or common (1) categories.

`sil_width` corresponds to the silhouette widths, which measure how well each species matches its own cluster (Eq. 2 Supporting Information). Silhouette values can range between -1 (the species fits the opposite cluster perfectly) and $+1$ (the species fits its own cluster perfectly) (Kaufman & Rousseeuw, 1990). A bar graph of silhouette widths gives an overview of the species allocation in FQAnts (Fig. 1b). The high positive values of most of the rare ant species indicate that they are well matched to its own cluster. Common ant species show smaller silhouette widths, suggesting a weaker cluster. In particular, the negative silhouette of species 27 indicates a poor fit to the common-species group (Fig. 1b), which conforms to its position in the AO plot (Fig. 1a).

`Common.I` represents the C_i s, i.e. the probability of each species being common. (Given that $C_i = 1 - R_i$, `fuzzyq` only returns C_i s). FuzzyQ includes a bootstrap procedure to compute the confidence intervals of C_i s. First, `fuzzyqBoot` generates and applies `fuzzyq` to N bootstrap replicates by site of the species abundance matrix:

```
> BS.FQAnts <- fuzzyqBoot(antsA, N= 1e3, level='spp')
```

Then `fuzzyqCI` computes the confidence intervals (95% by default) of C_i s based on these replicates:

```
> BS.FQAnts <- fuzzyqCI (BS.FQAnts, fq = FQAnts, method="bca")
```

A plot of the C_i s and their confidence intervals inform us of the nature of the ant species community in terms of their distribution along the rarity-commonness gradient (Fig. 1c).

(3) `FQAnts$global` returns the community-level metrics:

```
> FQAnts$global
silw.rar    silw.com    silw.all    commI.rar    commI.com    N.Dunn
0.87465334  0.32063735  0.79034656  0.08052098  0.77557146  0.69466718
```

`silw.rar` and `silw.com` (hereafter $\bar{S}_C$ and $\bar{S}_R$) are the average silhouettes' widths of the common and rare species that measure the coherence of the common- and rare-species clusters. Cluster coherence can be also measured with the corresponding average commonness coefficients `commI.rar` and `commI.com` (hereafter $\bar{C}_C$ and $\bar{C}_R$). `silw.all` and `N.Dunn` represent the average silhouette width of the whole community ($\bar{S}$) and the normalized Dunn's partition coefficient (D')

(Eq. 3 Supporting Information), respectively, which are metrics of the strength of overall classification (Kaufman & Rousseeuw, 1990).

We can also build confidence intervals for these parameters with `fuzzyqBoot` and `fuzzyqCI`:

```
> BS.global <- fuzzyqBoot (antsA, Nboot, level='global')
> BS.global <- fuzzyqCI (BS.global, fq = FQAnts, method="bca")
```

Let us compare the results of common and rare average silhouette values:

```
> BS.global[, 1:2]
      silw.rar  silw.com
Lower 0.8541929 0.1638335
Upper 0.8897733 0.5340427
```

Rare species have a higher average silhouette width and narrower 95% confidence interval, which indicates that rare species form a more coherent cluster than common ones.

3 WORKED-OUT EXAMPLES

We demonstrate the new method and its capabilities, with two real datasets involving the comparison of related and unrelated communities, respectively. When comparing several communities, one must consider how to deal with species absences. Absences may be due to (a) eco-evolutionary constraints, e.g. impassable geographical barriers (structural absence), (b) sampling variability, i.e. the species is indeed present but went undetected in the survey (random absence), or (c) methodological errors, e.g. misidentifications or low resolution to discriminate among species (false absences) (Blasco-Moreno, Pérez-Casany, Puig, Morante, & Castells, 2019). Although `fuzzyq` cannot deal with (c), it would produce different metrics in (a) and (b) situations and researchers should make an informed decision based on the nature of their system. The logical argument `rm.absent` in `fuzzyq` specifies whether species absences are to be treated as structural (TRUE; they are then removed from the data) or random (FALSE; the data is unchanged).

3.1 Mammal Data from Powdermill Biological Station 1979-1999

To illustrate how to monitor changes in species commonness in a community, we used a long-term (1979-1999) time series of small mammal abundances from the Powdermill Biological Station in Pennsylvania, USA (Merritt, 2018). Mammals were captured in a 1-ha live trapping grid consisting of 10×10 quadrats of trap stations at 10-m intervals. Trapping was conducted bimonthly from September 1979 to October 1999. For the sake of demonstration, the abundance of each mammal species was aggregated per quadrat and per year in order to capture the annual

variation in commonness of each species and we assumed that the pool of species did not change over the study period (random absences).

The C_i s indicated that two and seven of the 14 species could be categorized consistently as common and rare, respectively, throughout the study period. The C_i s of the remaining five varied considerably over the years (Fig. 2a). We modelled the change in commonness of a species in the latter group, *Glaucomys volans* (GV), by fitting a Generalized Additive Model. We used **fuzzyqBoot** to generate bootstrap replicates to fit 95% confidence intervals to the model. The fitted model was also used to predict the change in C_i five years ahead. (Details on model fitting are given in an accompanying R script. See Data Availability below.) The model suggests a progressive increase in C_i of *G. volans* over the years and predicts a similar increase rate in the following years (Fig. 2b).

3.2. Helminth communities of the so-iuy mullet in native and introduced areas

We compared the patterns of commonness and rarity of helminth communities of the so-iuy mullet *Planiliza haematocheilus* in its native (Sea of Japan) and introduced (Sea of Azov and Black Sea) areas (Llopis-Belenguer, Blasco-Costa, Balbuena, Sarabeev, & Stouffer, 2020). We used here 12 and 7 surveys in the introduced and native areas, respectively, in which the number of fish sampled was ≥ 20 , totalling 378 and 192 fish, respectively. Based on biogeographical evidence, species absences within the native and introduced areas were treated as random ones (Kostadinova, 2008).

We used **fuzzyq** to compute $\bar{S}$, $\bar{S}_C$, $\bar{S}_R$, $\bar{C}_C$, $\bar{C}_R$ and D' of each survey. Differences in these metrics between surveys of the native and introduced areas were evaluated by Mann-Whitney tests (Fig. 3). In the introduced area, $\bar{S}$ and D' were significantly higher (Fig. 3a, b), indicating a clearer distinction between common and rare species in the introduced area than in the native one. Likewise, rare species had significantly higher $\bar{S}_R$ and lower $\bar{C}_R$ in the introduced area (Fig. 3e, f). By contrast, there was no evidence for significant differences in mean values between areas in $\bar{S}_C$ and $\bar{C}_C$ (Fig. 3c, d). These results conform to previous work that indicates that the introduction of the mullet so-iuy in the new area entailed a deep structural change in its helminth communities (Sarabeev, Balbuena, & Morand, 2017; Llopis-Belenguer et al., 2020). Most native species were lost and the majority of species in the introduced area were acquired from local grey mullet species (see Figs. S1, S2 and further details in Supporting Information). Since newly acquired parasite species are expected to lack specific adaptations to the new host, this would account for

their pronounced rarity compared to rare species in the native area (Sarabeev, Balbuena, & Morand, 2018).

4 EFFECT OF NUMBER OF SITES

A key question for users interested in comparing different communities is whether the community-level estimates depend on the number of sites sampled. We examined this issue using 20 datasets compiled in Calatayud et al. (2019) and Jeliaskov et al. (2020) involving 87+ sites. These included communities varying widely in taxonomic composition, species richness, spatial scale and abundance currency used. (See Table S1 in Supporting Information for details.) Being N the total number of sites of a given dataset, global metrics ($\bar{S}$, $\bar{S}_C$, $\bar{S}_R$, $\bar{C}_C$, $\bar{C}_R$ and D^*) were computed in most cases dropping successively 1, 2, 3, ..., $N-10$ sites randomly drawn (without replacement) from the dataset. (In species-poor communities or communities with very sparsely distributed species, the series was 1, 2, 3, ..., $N-20$). Species absences in each draw were treated as structural, because our goal was to examine how incomplete species coverage resulting from low sample sizes affected the estimation of global metrics. (See also a similar experiment related to the ability of detect species in Supplementary Information).

Fig. 4 shows the variation of the global metrics with the number of sites in four datasets. (Results for the remaining 16 datasets are given in Fig. S3, Supporting Information.) No clear trend of variation with the number of sites was apparent (Fig. 4 and Fig. S3, Supporting Information). Although in some datasets large fluctuations occurred ($\bar{S}_C$ in particular was quite labile in some examples), the results suggest that 30 to 50 sites are sufficient to yield reliable estimates. We also recommend building bootstrap confidence intervals to assess their accuracy.

5 FINAL REMARKS

FuzzyQ provides a new quantitative framework to study the distribution of common and rare species in ecological communities. The approach supplies simple and intuitive ecological indicators that can give both clear, actionable insights into the nature of ecological communities and a powerful way to monitor quantitatively environmental change on ecosystems. We show that FuzzyQ works satisfactorily with a wide range of communities varying in species richness, dispersion and abundance currencies. The only obvious limitation in the application of FuzzyQ is that fuzzy clustering requires that the number of clusters $k \geq n/2 - 1$, where n is the number of observations (species in our case) (Kaufman & Rousseeuw, 1990). So FuzzyQ cannot be applied to communities composed of ≤ 5 species. In addition, the application of `fuzzyqBoot` in

communities with low number of species can lead to a number of null replicates because of this limitation.

The fuzzy clustering approach is versatile as it can be readily adapted to other categorizations (by considering more clusters) or to other criteria of rarity (by introducing additional/different traits when computing the species dissimilarity matrix). For instance, Rabinowitz (1981) distinguished seven different forms of rarity, associated with quite different risks of extinction, based on dividing three continuous variables into respective binary categories. Since opinions differ on where and how to place the dividing thresholds (Reed et al., 2019), fuzzy clustering could provide a more objective categorization tool and a way to assess how well species fit in their assigned category.

Comparison among communities is greatly facilitated by the fact that FuzzyQ is relatively independent of the number of sites or sampling units considered. However, its use in comparative settings comes with an important caveat. Since FuzzyQ is based on AOs and occupancy is known to vary with spatial scale (Hui, Veldtman, & McGeoch, 2010; Steenweg, Hebblewhite, Whittington, Lukacs, & McKelvey, 2018), differences in scaling can compromise comparison among communities. In our second working example, helminth communities of individual fish were evaluated as sites. Therefore, we consider that the comparison makes biological sense. However, we cannot completely rule out that potential differences between the native and introduced areas in fish mobility could introduce a hidden bias (Steenweg et al., 2018).

Likewise, it has been shown that rarity at coarse scales can be substantially biased because species of similar occupancies at that level may have very different occupancies at finer scales (He & Condit, 2007). Thus, assessment and monitoring of rarity should be performed at the appropriate scale for suitable conservation and management plans. Nevertheless, for samples taken at nested spatial or temporal scales, FuzzyQ provides a convenient tool to assess how scale affects patterns of commonness and rarity. Therefore, we consider that FuzzyQ is a potentially valuable analytical tool in community ecology and conservation biology.

ACKNOWLEDGEMENTS

We thank Xavier Arnan and co-workers for allowing us to incorporate their ant species dataset to the FuzzyQ package. We are indebted to the anonymous reviewers and associate editor for providing insightful comments and suggestions for additional work which have resulted in a much improved paper. Study funded by the Ministry of Science and Innovation, Spain (PID2019-104908GB-I00). VLS was supported by the Polish National Agency for Academic Exchange

(PPN/ULM/2019/1/00177/U/00001). The mammal Powdermill data was obtained with the support of NSF Grants BSR-8702333-06, DEB-9211772, DEB-9411974, DEB-0080381 and DEB-0621014.

AUTHORS' CONTRIBUTIONS

JAB conceived the idea. JAB and SM developed the theory and outlined the study. CMB, CLB and JAB wrote the scripts and developed the package. CLB, IBC, VLS and SM set and verified the analytical methods. All authors contributed decisively to shape the research, provided critical feedback on the drafts and gave final approval for submission.

DATA AVAILABILITY

Table 1 provides information of the availability of the datasets used herein. R scripts and R markdown files to run the illustrative examples, and the FuzzyQ R package are available on Github, versioned and archived on Zenodo (Balbuena et al., 2021).

References

- Alroy, J. (2015). The shape of terrestrial abundance distributions. *Science Advances*, *1*(8), e1500082. <https://doi.org/10.1126/sciadv.1500082>
- Antão, L. H., Connolly, S. R., Magurran, A. E., Soares, A., & Dornelas, M. (2017). Prevalence of multimodal species abundance distributions is linked to spatial and taxonomic breadth. *Global Ecology and Biogeography*, *26*(2), 203–215. <https://doi.org/10.1111/geb.12532>
- Arnan, X., Gaucherel, C., & Andersen, A. N. (2011). Dominance and species co-occurrence in highly diverse ant communities: a test of the interstitial hypothesis and discovery of a three-tiered competition cascade. *Oecologia*, *166*(3), 783–794. <https://doi.org/10.1007/s00442-011-1919-y>
- Bagnaro, A., Baltar, F., Brownstein, G., Lee, W. G., Morales, S. E., Pritchard, D. W., & Hepburn, C. D. (2020). Reducing the arbitrary: fuzzy detection of microbial ecotones and ecosystems – focus on the pelagic environment. *Environmental Microbiome*, *15*(1), 16. <https://doi.org/10.1186/s40793-020-00363-w>
- Balbuena, J.A., Monlleó-Borrull, C.; Llopis-Belenguer, C.; Blasco-Costa, I., Sarabeev, V.L., & Morand, S. (2021). Code for Fuzzy Quantification of Common and Rare Species in Ecological Communities (Version v1.0). *Zenodo*. <http://doi.org/10.5281/zenodo.4469291>
- Barbosa, A. M. (2015). fuzzySim: applying fuzzy logic to binary similarity indices in ecology. *Methods in Ecology and Evolution*, *6*(7), 853–858. <https://doi.org/10.1111/2041-210x.12372>

- Blasco-Moreno, A., Pérez-Casany, M., Puig, P., Morante, M., & Castells, E. (2019). What does a zero mean? Understanding false, random and structural zeros in ecology. *Methods in Ecology and Evolution*, 10(7), 949–959. <https://doi.org/10.1111/2041-210X.13185>
- Calatayud, J., Andivia, E., Escudero, A., Melián, C. J., Bernardo-Madrid, R., Stoffel, M., ... Madrigal-González, J. (2019). Positive associations among rare species and their persistence in ecological assemblages. *Nature Ecology & Evolution*, 4(1), 40–45. <https://doi.org/10.1038/s41559-019-1053-5>
- Davies, K. F., Margules, C. R., & Lawrence, J. F. (2000). Which traits of species predict population declines in experimental forest fragments? *Ecology*, 81(5), 1450–1461. [https://doi.org/10.1890/0012-9658\(2000\)081\[1450:WTOSPP\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081[1450:WTOSPP]2.0.CO;2)
- Dee, L. E., Cowles, J., Isbell, F., Pau, S., Gaines, S. D., & Reich, P. B. (2019). When do ecosystem services depend on rare species? *Trends in Ecology & Evolution*, 34(8), 746–758. <https://doi.org/10.1016/j.tree.2019.03.010>
- Dick, D. G., & Laflamme, M. (2018). Fuzzy ecospace modelling. *Methods in Ecology and Evolution*, 9(6), 1442–1452. <https://doi.org/10.1111/2041-210x.13010>
- Efron, B., & Tibshirani, R. J. (1994). An introduction to the bootstrap. New York, NY: Chapman and Hall. <https://doi.org/10.1201/9780429246593>
- Enquist, B. J., Feng, X., Boyle, B., Maitner, B., Newman, E. A., Jørgensen, P. M., ... McGill, B. J. (2019). The commonness of rarity: Global and future distribution of rarity across land plants. *Science Advances*, 5(11), eaaz0414. <https://doi.org/10.1126/sciadv.aaz0414>
- Fiorentino, D., Pesch, R., Guenther, C., Gutow, L., Holstein, J., Dannheim, J., ... Wiltshire, K. (2017). A ‘fuzzy clustering’ approach to conceptual confusion: how to classify natural ecological associations. *Marine Ecology Progress Series*, 584, 17–30. <https://doi.org/10.3354/meps12354>
- Flather, C. H., & Sieg, C. H. (2007). Species rarity: definition, causes, and classification. In M.G. Raphael & R. Molina (Eds.), *Conservation of rare or little-known species: biological, social, and economic considerations* (pp. 40–66). Washington D.C.: Island Press.
- Gaston, K. J. (1994). *Rarity* (1st ed., Vol. 13). Dordrecht, Netherlands: Springer. <https://doi.org/10.1007/978-94-011-0701-3>
- Gaston, K. J. (1997). What is rarity? In W.E. Kunin & K.J. Gaston (Eds.), *The biology of rarity. Causes and consequences of rare-common differences* (pp. 30-47). Dordrecht, Netherlands: Springer. https://doi.org/10.1007/978-94-011-5874-9_3

- Gaston, K. J., Blackburn, T. M., Greenwood, J. J. D., Gregory, R. D., Quinn, R. M., & Lawton, J. H. (2000). Abundance–occupancy relationships. *Journal of Applied Ecology*, 37(s1), 39–59. <https://doi.org/10.1046/j.1365-2664.2000.00485.x>
- Gaston, K. J., & Fuller, R. A. (2008). Commonness, population depletion and conservation biology. *Trends in Ecology & Evolution*, 23(1), 14–19. <https://doi.org/10.1016/j.tree.2007.11.001>
- Gaston, K. J., & He, F. (2011). Species occurrence and occupancy. In A.E. Magurran & B.J. McGill (Eds.), *Biological diversity: frontiers in measurement and assessment* (pp. 141–151). Oxford, U.K.: Oxford University Press
- Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27(4), 857–871. <https://doi.org/10.2307/2528823>
- Gray, J. S., Bjørgesæter, A., & Ugland, K. I. (2005). The impact of rare species on natural assemblages. *Journal of Animal Ecology*, 74(6), 1131–1139. <https://doi.org/10.1111/j.1365-2656.2005.01011.x>
- Hanski, I. (1991). Single-species metapopulation dynamics: concepts, models and observations. *Biological Journal of the Linnean Society*, 42(1–2), 17–38. <https://doi.org/10.1111/j.1095-8312.1991.tb00549.x>
- He, F., & Condit, R. (2007). The distribution of species: occupancy, scale, and rarity. In D. Storch, P. Marquet, & J. Brown (Eds.), *Scaling Biodiversity* (Ecological Reviews, pp. 32–50). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511814938.005>
- Hui, C., Veldtman, R., & McGeoch, M. A. (2010). Measures, perceptions and scaling patterns of aggregated species distributions. *Ecography*, 33(1), 95–102. <https://doi.org/10.1111/j.1600-0587.2009.05997.x>
- Jeliazkov, A., Mijatovic, D., Chantepie, S., Andrew, N., Arlettaz, R., Barbaro, L., ... Chase, J. M. (2020). A global database for metacommunity ecology, integrating species, traits, environment and space. *Scientific Data*, 7, 6. <https://doi.org/10.1038/s41597-019-0344-7>
- Kaufman, L., & Rousseeuw, P. J. (1990). Finding groups in data: An introduction to cluster analysis. New York, U.S.A.: John Wiley & Sons. <https://doi.org/10.1002/9780470316801>
- Kostadinova, A. (2008). A checklist of macroparasites of *Liza haematocheila* (Temminck & Schlegel) (Teleostei: Mugilidae). *Parasites & Vectors*, 1(1), 48. <https://doi.org/10.1186/1756-3305-1-48>

- Leitão, R. P., Zuanon, J., Villéger, S., Williams, S. E., Baraloto, C., Fortunel, C., ... Mouillot, D. (2016). Rare species contribute disproportionately to the functional structure of species assemblages. *Proceedings of the Royal Society B: Biological Sciences*, 283(1828), 20160084. <https://doi.org/10.1098/rspb.2016.0084>
- Llopis-Belenguer, C., Blasco-Costa, I., Balbuena, J. A., Sarabeev, V., & Stouffer, D. B. (2020). Native and invasive hosts play different roles in host–parasite networks. *Ecography*, 43(4), 559–568. <https://doi.org/10.1111/ecog.04963>
- Loiseau, N., Mouquet, N., Casajus, N., Grenié, M., Guéguen, M., Maitner, B., ... Violle, C. (2020). Global distribution and conservation status of ecologically rare mammal and bird species. *Nature Communications*, 11, 5071. <https://doi.org/10.1038/s41467-020-18779-w>
- Magurran, A. E., & Henderson, P. A. (2011). Commonness and rarity. In A.E. Magurran & B.J. McGill (Eds.), *Biological diversity: frontiers in measurement and assessment* (pp. 97–104). Oxford, U.K.: Oxford University Press.
- McGill, B. (2011). Species abundance distributions. In A.E. Magurran & B.J. McGill (Eds.), *Biological diversity: frontiers in measurement and assessment* (pp. 105–122). Oxford, U.K.: Oxford University Press.
- McGill, B. J., Rampal S. Etienne, John S. Gray, David Alonso, Marti J. Anderson, Habtamu Kassa Benecha, ... Ethan P. White. (2007). Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, 10(10), 995–1015. <https://doi.org/10.1111/j.1461-0248.2007.01094.x>
- Merritt, J. (2018). Long term mammal data from Powdermill biological station 1979-1999 [Data set]. Environmental Data Initiative. <https://doi.org/10.6073/pasta/101d5d3dec9c688a7feca3ab2e969369>
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. Accessed 7 August 2020
- Rabinowitz, D. (1981). Seven forms of rarity. In: H. Synge (Ed.) *The biological aspects of rare plant conservation* (pp 205–217). Somerset, U.S.A.: John Wiley & Sons.
- Reed, J. M., Nguyen, A., Owens, A. C. S., & Lewis, S. M. (2019). Linking the seven forms of rarity to extinction threats and risk factors: an assessment of North American fireflies. *Biodiversity and Conservation*, 29(1), 57–75. <https://doi.org/10.1007/s10531-019-01869-7>
- Sarabeev, V., Balbuena, J. A., & Morand, S. (2017). Testing the enemy release hypothesis: Abundance and distribution patterns of helminth communities in grey mullets (Teleostei:

- Mugilidae) reveal the success of invasive species. *International Journal for Parasitology*, 47(10–11), 687–696. <https://doi.org/10.1016/j.ijpara.2017.05.006>
- Sarabeev, V., Balbuena, J. A., & Morand, S. (2018). Invasive parasites are detectable by their abundance-occupancy relationships: the case of helminths from *Liza haematocheilus* (Teleostei: Mugilidae). *International Journal for Parasitology*, 48(9–10), 793–803. <https://doi.org/10.1016/j.ijpara.2018.04.001>
- Steenweg, R., Hebblewhite, M., Whittington, J., Lukacs, P., & McKelvey, K. (2018). Sampling scales define occupancy and underlying occupancy–abundance relationships in animals. *Ecology*, 99(1), 172–183. <https://doi.org/10.1002/ecy.2054>
- Violle, C., Thuiller, W., Mouquet, N., Munoz, F., Kraft, N. J. B., Cadotte, M. W., ... Mouillot, D. (2017). Functional rarity: The ecology of outliers. *Trends in Ecology & Evolution*, 32(5), 356–367. <https://doi.org/10.1016/j.tree.2017.02.002>
- Werner, E. E., Davis, C. J., Skelly, D. K., Relyea, R. A., Benard, M. F., & McCauley, S. J. (2014). Cross-scale interactions and the distribution-abundance relationship. *PLOS ONE*, 9(5), e97387. <https://doi.org/10.1371/journal.pone.0097387>
- Williamson, M., & Gaston, K. J. (2005). The lognormal distribution is not an appropriate null hypothesis for the species–abundance distribution. *Journal of Animal Ecology*, 74(3), 409–422. <https://doi.org/10.1111/j.1365-2656.2005.00936.x>
- Yu, J., & Dobson, F. S. (2000). Seven forms of rarity in mammals. *Journal of Biogeography*, 27(1), 131–139. <https://doi.org/10.1046/j.1365-2699.2000.00366.x>

FIGURE CAPTIONS

FIGURE 1 Fuzzy quantification of common and rare species in a community of 46 ant species in 100 plots (Arnan et al., 2011, Calatayud et al., 2019). (a) Abundance occupancy relationship of species. The arrow points to the position of Species 27. (b) Silhouette bar plot of the species. The negative value of Species 27 suggests a poor fit to the cluster of common species. (c) Commonness indices of species. Error bars represent bias-corrected and accelerated 95% confidence intervals (Efron & Tibshirani, 1994) computed with 1,000 bootstrap samples by row of the abundance matrix. The horizontal line marks the 0.5 threshold separating rare and common species. Light blue and light red denote common and rare species, respectively.

FIGURE 2 Fuzzy quantification of common and rare species in a community of 14 mammal species sampled at the Powdermill Biological Station from 1979 to 1999 (Merritt, 2018). (a)

Variation of commonness indices in the study period. (b) Generalized Additive Model describing the variation in commonness of *Glaucomys volans* (GV) in the study period and predicted change (2001-2004). Blue points: observed values. Thick red line: fitted and predicted model (continuous and stippled lines). Thin red line: 95% confidence interval of the model. Stippled orange line: 0.5 threshold between rare and common species. Species abbreviations: **BB**, *Blarina brevicauda*; **CG**, *Clethrionomys gapperi*; **DV**, *Didelphis virginiana*; **GV**, *Glaucomys volans*; **MF**, *Mustela frenata*; **MM**, *Marmota monax*; **NI**, *Napaeozapus insignis*; **PL**, *Peromyscus leucopus*; **PM**, *Peromyscus maniculatus*; **SC**, *Sorex cinereus*; **SD**, *Sorex dispar*; **SF**, *Sorex fumeus*; **SH**, *Sorex hoyi*; **TS**, *Tamias striatus*.

FIGURE 3 Community-level estimates (points) and median values (horizontal lines) measuring the strength of fuzzy clustering of common and rare species in helminth communities of *Planiliza haematocheilus* of 7 and 12 surveys in the host's native (Japan Sea) and introduced (Azov and Black Seas) areas. Results of Mann-Whitney tests testing differences between areas are also displayed. (a) Average silhouette widths of all species; (b) normalized Dunn's coefficient; (c) average silhouette widths and (d) average commonness indices of common species; (e) average silhouette widths and (f) average commonness indices of rare species.

FIGURE 4 Variation of community-level metrics with number of sites in four, 87+-site, databases from Jeliaskov et al. (2020) and Calatayud et al. (2019). (a) Barbaro2009b; (b) Chmura2016; (c) ants_data_Xavi_Darwin_C; (d) Goncalves2014a. Abbreviations: $\bar{S}_R$, average silhouette rare species; $\bar{S}_C$, idem common species; $\bar{S}$, idem all species; $\bar{C}_R$, Commonness coefficient rare species; $\bar{C}_C$, idem common species; D' , Normalized Dunn's coefficient. Details of these datasets are given in Table S2, Supporting Information.

TABLE 1 Availability of the datasets used with FuzzyQ.

Datasets	Source	Reference
Mammals	https://doi.org/10.6073/pasta/101d5d3dec9c688a7feca3ab2e969369	Merritt (2018)
Powdermill		
Helminths so-iuy mullet	https://doi.org/10.7910/DVN/IWIKOL	Llopis- Belenguer et al. (2020)
Large communities	https://idata.idiv.de/ddm/Data/ShowData/286	Jeliazkov et al. (2020)
Large communities	https://doi.org/10.6084/m9.figshare.9906092	Calatayud et al. (2019)







