

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

Juan A. Balbuena<sup>1,\*</sup> Clara Montlleó<sup>1</sup> Cristina Llopis-Belenguer<sup>1</sup>  
Isabel Blasco-Costa<sup>2</sup> Volodimir L. Sarabeev<sup>3</sup> Serge Morand<sup>4,5</sup>

26 July 2020

<sup>1</sup> *Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, Valencia, Spain*

<sup>2</sup> *Department of Invertebrates, Natural History Museum of Geneva, Geneva, Switzerland*

<sup>3</sup> *Department of Biology, Zaporizhzhia National University, Zaporizhzhia, Ukraine*

<sup>4</sup> *Institut des Sciences de l'Evolution, CNRS, CC065, Université de Montpellier, Montpellier Cedex, France; CIRAD ASTRE*

<sup>5</sup> *Faculty of Veterinary Technology, Kasetsart University, 10400 Bangkok, Thailand*

\* Correspondence: Juan A. Balbuena <j.a.balbuena@uv.es>

## FuzzyQ – Ant species example

We illustrate the use and capabilities of FuzzyQ with a dataset of ant species (ants\_Darwin\_A from Calatayud et al. (2019)). The dataset collates the abundance of 46 species in 100, 18 × 18 m plots sampled in the Northern Territory, Australia (Arnan et al., 2011).

Function `fuzzyq` performs a fuzzy clustering algorithm (Kaufman & Rousseeuw, 1990) that evaluates all pairwise dissimilarities among species in the dataset. Based on its Area Occupancy (AO) (i.e., the mean abundance and number of sites occupied), each species is allocated into two a-priori defined clusters of common and rare species.

```
library(FuzzyQ)
data(antsA)      # loads the ant database (site-by-species abundance matrix)

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

Note that we set the argument `sorting = TRUE` to sort the species by cluster. This facilitates effective visualization of the rare and common species composition of the community. `FQAnts$A_0`, returns the AO information of each species. We use `A0plot` to show the AO relationship of the ant community (Fig. 1).

```
col.RC <- c("brown2", "turquoise3")
A0plot(FQAnts, col.RC, pch = 16)
legend(0, 10.5, c("Rare", "Common"), col = col.RC, pch = 16)
```

The allocation of species to clusters is stored in `FQAnts$spp`. Each species has been allocated to the rare (cluster = 0) or common cluster (cluster = 0 and 1, respectively):

```
##      cluster  sil_width  Common.I
## sps_15      0  0.49317392 0.45193394
## sps_8       0  0.51278584 0.43707371
```

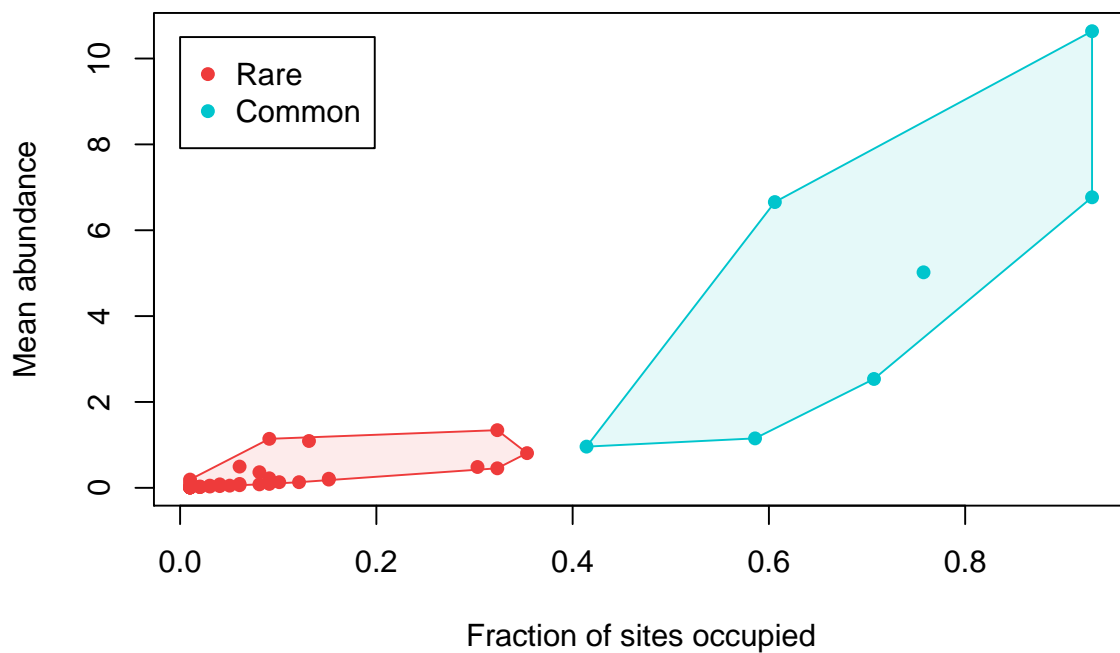


Figure 1: Relationship between mean abundance and number of sites occupied of 46 ant species

## sps_45	0	0.62327350	0.36020316
## sps_35	0	0.65201657	0.33660336
## sps_18	0	0.80403066	0.17971997
## sps_19	0	0.83367067	0.14262525
## sps_29	0	0.86173265	0.11387856
## sps_36	0	0.86290322	0.11228338
## sps_23	0	0.89117336	0.07413132
## sps_10	0	0.90167855	0.05756174
## sps_20	0	0.90370956	0.05596081
## sps_5	0	0.90547644	0.05376546
## sps_14	0	0.90781862	0.05050306
## sps_9	0	0.91221695	0.04389605
## sps_4	0	0.91295943	0.04268709
## sps_39	0	0.91729545	0.03593477
## sps_25	0	0.91754402	0.03265451
## sps_1	0	0.92007954	0.02844395
## sps_13	0	0.92007954	0.02844395
## sps_26	0	0.92007954	0.02844395
## sps_28	0	0.92007954	0.02844395
## sps_40	0	0.92007954	0.02844395
## sps_41	0	0.92007954	0.02844395
## sps_43	0	0.92007954	0.02844395
## sps_44	0	0.92007954	0.02844395
## sps_34	0	0.92036413	0.02840823
## sps_22	0	0.92049340	0.02788790
## sps_12	0	0.92329475	0.02630374
## sps_46	0	0.92350710	0.02586522
## sps_7	0	0.92387440	0.02340876

```
## sps_11      0  0.92387440 0.02340876
## sps_17      0  0.92387440 0.02340876
## sps_6       0  0.92412791 0.02307659
## sps_30      0  0.92497786 0.02319410
## sps_37      0  0.92562230 0.02129898
## sps_2       0  0.92579517 0.02108267
## sps_42      0  0.92583965 0.02144117
## sps_24      0  0.92585524 0.02157493
## sps_21      0  0.92588366 0.02099071
## sps_27      1 -0.44650248 0.52553230
## sps_3       1  0.03065189 0.71356545
## sps_38      1  0.41606315 0.83454643
## sps_31      1  0.51494510 0.76967199
## sps_16      1  0.54938659 0.86063493
## sps_33      1  0.58273755 0.88190256
## sps_32      1  0.59717967 0.84314656
```

The commonness indices (`Common.I`) displayed above can be interpreted as the probability of each species being common given the AO observed. We wish to plot these indices with their 95% confidence intervals. To this end, `fuzzyqBoot` generates 1,000 bootstrap replicates by site and `fuzzyqCI` estimates the confidence intervals based on the replicates:

```
Nboot = 1e3
set.seed = 2020
BS.FQAnts <- fuzzyqBoot(antsA, Nboot, level='spp')
BS.FQAnts <- fuzzyqCI(BS.FQAnts, fq=FQAnts, method="bca")
```

In this case we chose the `bca` (bias corrected and accelerated) method. Two additional methods, bias corrected (`method = "bc"`) and percentile (`method = "pct"`) are implemented in `fuzzyqCI`. Prior to plotting, the confidence intervals need to be sorted as per the clustering produced by `FQAnts`:

```
BS.FQAnts <- sortClus(BS.FQAnts, FQAnts)
```

Figure 2 shows the commonness indices and confidence intervals of each species in our dataset.

```
spp <- FQAnts$spp
plot(spp[,3], cex.axis=0.8, xaxt='n', ylab="Commonness index",
     ylim=c(0, max(BS.FQAnts)), xlab="Species", col=col.RC[spp[,1]+1],
     pch=16, cex=0.8, las = 1)
ebar.int <- seq_len(nrow(spp)) # use arrows function to draw CIs:
arrows(ebar.int, BS.FQAnts["Lower", ], ebar.int, BS.FQAnts["Upper", ],
       length= 0, col=col.RC[spp[,1]+1])
abline(h=0.5, col="orange3", xpd=FALSE)
axis(1, at=ebar.int, labels=row.names(spp), las=2, cex.axis=0.6)
```

`FQAnts$spp` also provides the silhouette value of each species (`sil_width`). Silhouettes are measures of how similar abundance and occupancy of each species are to its own cluster relative to species in the other cluster. They can be displayed as a barplot that informs us about the nature of the community in terms of the distribution of common and rare species (Fig. 3).

```
barplot(spp[,2], names.arg=row.names(spp),
        col=col.RC[spp[,1]+1],
        las=2, cex.names=0.6, xlab="Species", ylab="Silhouette width",
        cex.axis = 0.8)
legend(0, -0.1, c("Rare", "Common"), col = col.RC, pch = 15, cex = 0.8)
```

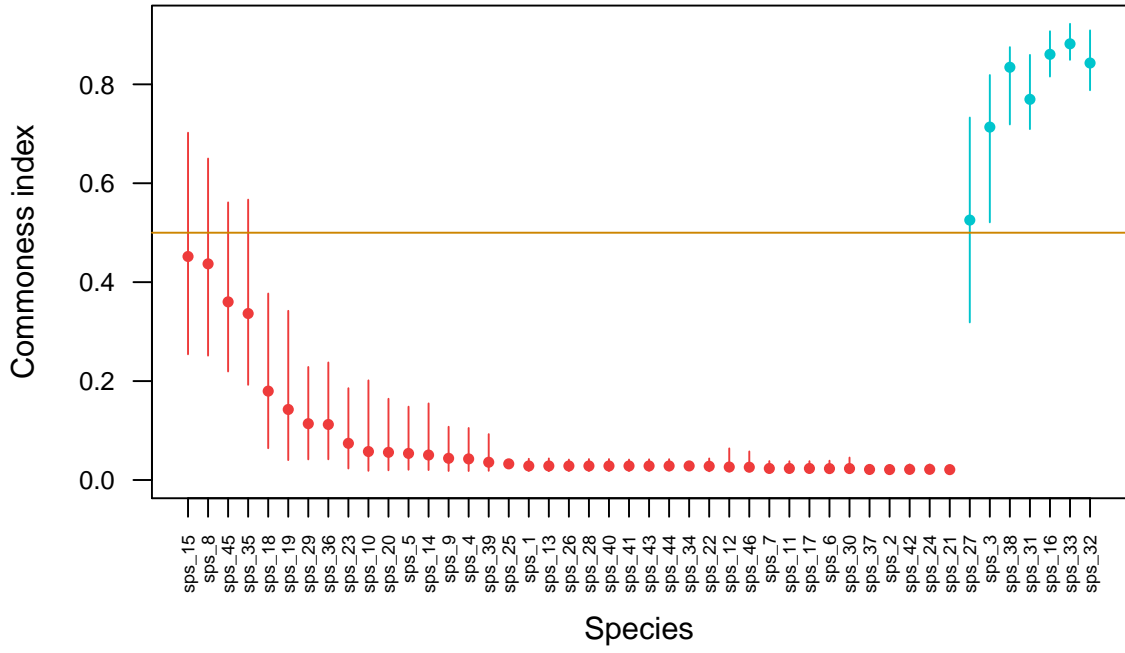


Figure 2: Commonness indices and 95% bootstrap confidence intervals of 46 ant species generated with a fuzzy clustering algorithm. The horizontal line indicates a commonness index of 0.5

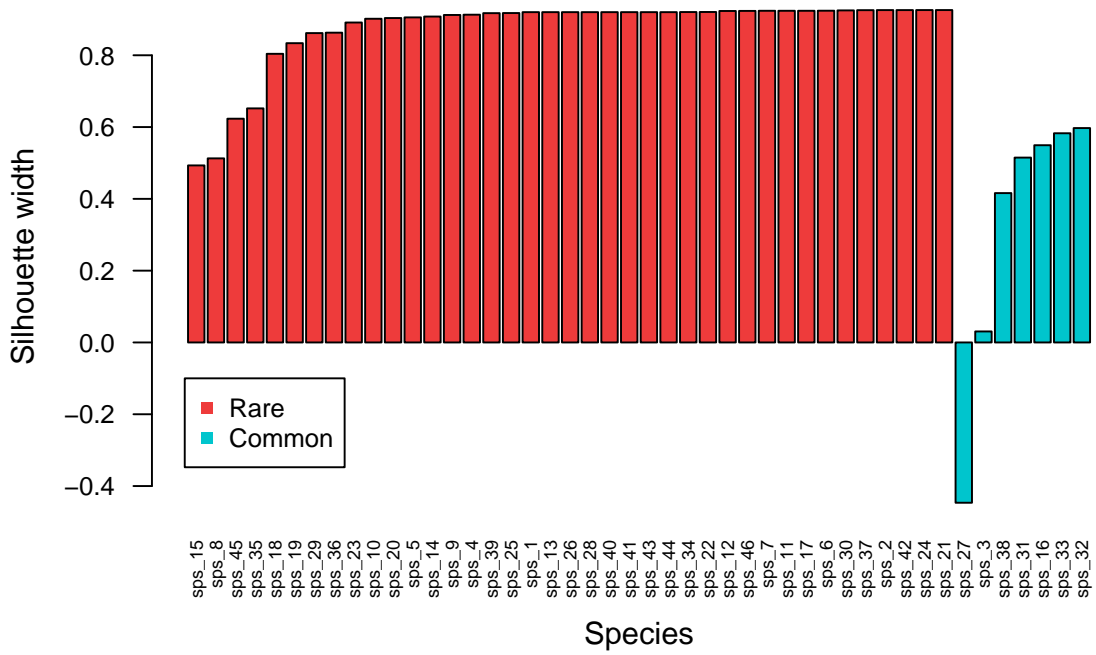


Figure 3: Silhouettes of 46 ant species generated with a fuzzy clustering algorithm

Common ant species show smaller silhouettes, suggesting a weaker cluster and the negative silhouette of species 27 indicates a poor fit to the common-species group (Fig. 3).

We can assess the strength of each cluster by computing the average silhouette widths. Alternatively, we can calculate the average commonness indices of rare and common species. In addition, the overall clustering structure of the community can be evaluated with the global silhouette average or the normalized Dunn's index (Kaufman & Rousseeuw, 1990). These parameters are to be found in `FQAnts$global`. To assess their variability, we use `fuzzyqBoot` to produce 1,000 bootstrap replicates by site (Fig. 4)

```
BS.global <- fuzzyqBoot (antsA, Nboot, level='global')

par(mar=c(2, 4.6, 1, 2))
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE), widths=c(2.75,1.25),
       heights=c(1,1))
boxplot.matrix(BS.global$fq.rep[,1:3], cex.axis=0.8,
               names=c("rare", "common", "global"),
               ylab="Average silhouette width",
               xlab="", col=c(col.RC,"chartreuse3"),
               pch=16, cex=0.4)
points(FQAnts$global[1:3], pch=21, col="darkred", bg="papayawhip")
mtext("a)", side=2, at=0.95, line=3.6, las=1, cex=1.1)
boxplot.matrix(BS.global$fq.rep[,4:5], cex.axis=0.8,
               ylab="Average commonness index", names=c("rare", "common"),
               xlab="", col=col.RC,
               pch=16, cex=0.4)
points(FQAnts$global[4:5], pch=21, col="darkred", bg="papayawhip")
mtext("b)", side=2, at=0.92, line=3.6, las=1, cex=1.1)
mtext("c)", side=2, at=0.92, line=-23, las=1, cex=1.1)
boxplot(BS.global$fq.rep[,6], las=2, cex.axis=0.8,
        ylab="Normalized Dunn's coefficient",
        xlab="", col= "chartreuse3",
        pch=16, cex=0.4)
points(FQAnts$global[6], pch=21, col="darkred", bg="papayawhip")
```

Both the average global silhouette width and the normalized Dunn index observed suggest a strong clustering structure separating common and rare species (Kaufman & Rousseeuw, 1990). Compared with common ones, rare species showed a higher average silhouette width, a much lower average commonness coefficient and less variation, suggesting that they form a harder cluster (Fig. 4).

## References

- 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. doi:10.1007/s00442-011-1919-y
- 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. doi:10.1038/s41559-019-1053-5
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc. doi:10.1002/9780470316801

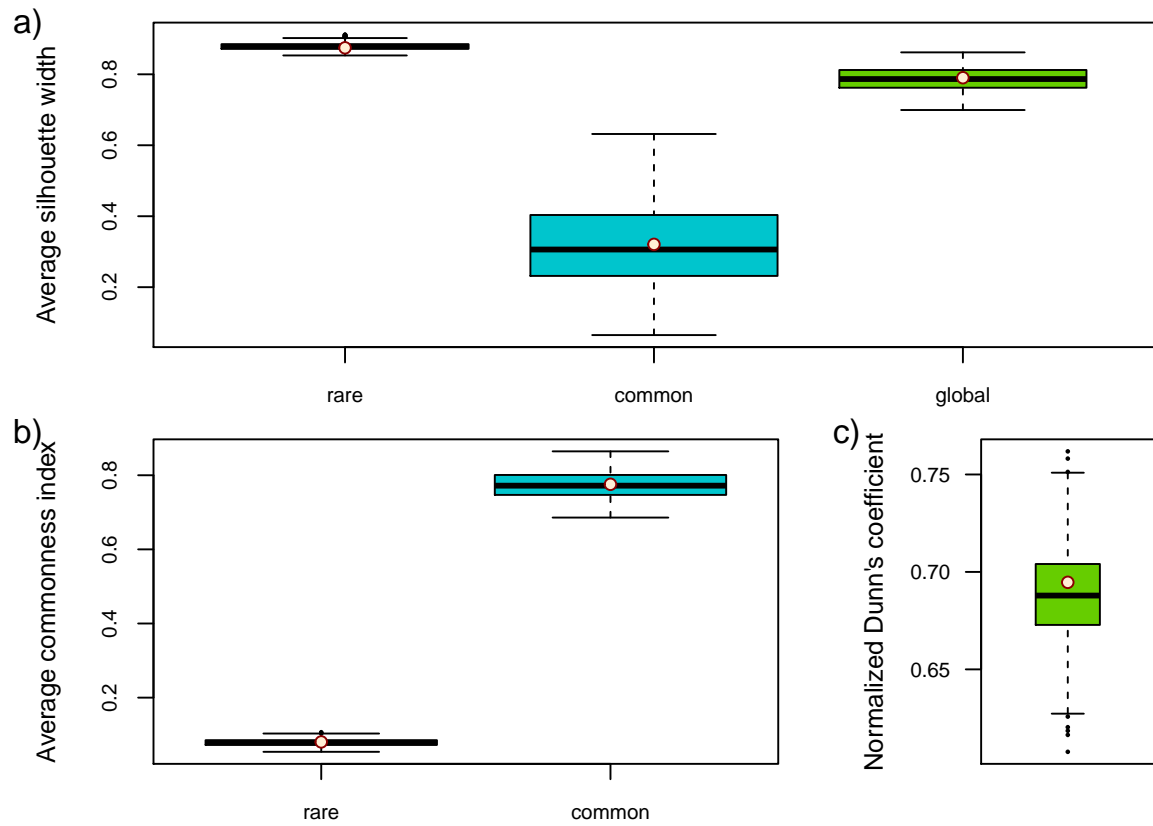


Figure 4: Global clustering parameters of a community of 46 ant species generated with a fuzzy clustering algorithm. Boxplots represent the variation of 1,000 bootstrap replicates. Points correspond to the estimates of the database. a) Average silhouette widths of rare and common species and globally; b) Average Commonness Indices of rare and common species; c) Normalized Dunn's index