

Biases and distribution patterns in hard-bodied microscopic animals (Acari: Halacaridae): Size does not matter, but generalism and sampling effort do

Iñigo Rubio-López^{1,2} | Fernando Pardos² | Diego Fontaneto¹ | Alejandro Martínez¹ | Guillermo García-Gómez^{2,3}

¹National Research Council of Italy (CNR), Water Research Institute (IRSA), Molecular Ecology Group (MEG), Verbania, Italy

²Department of Biodiversity, Ecology and Evolution, Faculty of Biology, University Complutense of Madrid, Madrid, Spain

³Department of Earth, Oceans and Ecological Sciences, School of Environmental Sciences, University of Liverpool, Liverpool, UK

Correspondence

Iñigo Rubio-López, National Research Council of Italy (CNR), Water Research Institute (IRSA), Molecular Ecology Group (MEG), Largo Tonolli 50, Verbania 28922, Italy.

Email: inirubio@ucm.es; inir725@gmail.com

Funding information

Erasmus Mobility

Editor: Chris Burridge

Abstract

Aim: The interplay between distribution ranges, species traits and sampling and taxonomic biases remains elusive amongst microscopic animals. This ignorance obscures our understanding of the diversity patterns of a major component of biodiversity. Here, we used marine Halacaridae to explore whether differences between marine provinces can explain their distribution patterns or if differential sampling efforts across regions prevent any macroecological inference. Furthermore, we test if certain functional traits influence their distribution patterns.

Location: Europe.

Results: Whereas geographical variables provided a better explanation for differences in species composition, sampling effort and distance from marine biological stations accounted for the majority of differences in European Halacaridae richness. Species occurring in more habitats showed broader geographical ranges and accumulated more records. Species traits like generalism affected the distribution of halacarid species.

Main Conclusions: We propose that the sampling effort of halacarid mites in Europe might be explained by two different cognitive biases: the convenience of selecting certain sampling localities compared to others and the tendency of zoologists to scrutinise habitats where their target organisms are more common.

KEYWORDS

biogeography, diversity, macroecology, marine meiofauna, meiobenthos, mites, sampling bias, spatial scales, species richness

1 | INTRODUCTION

Not every animal is equally easy to find, even for those biologists who have spent years studying specific groups. Whereas some

animals are seemingly everywhere (Bik et al., 2010; Danovaro & Gambi, 2022; Fontaneto et al., 2006), others are known only from a single locality (Foissner, 2006; Garrick et al., 2004; Ribera et al., 2002; Smith & Wilkinson, 2007). The distribution ranges of most species

Alejandro Martínez and Guillermo García-Gómez contributed equally to this work.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Diversity and Distributions* published by John Wiley & Sons Ltd.

vary amongst these two extremes. As a corollary, biological diversity is unevenly distributed throughout the Earth, defining patterns of species richness that change with spatial scale (Gaston, 2000; Hawkins et al., 2003; MacArthur & Wilson, 2016). These patterns reflect the interaction between environmental parameters and organisms' traits (Gotelli et al., 2009; Willig et al., 2003) but also the historical events that have affected the spatial units under consideration (Gaston, 2009; Schemske, 2002; Willig et al., 2003).

Concomitantly, our knowledge of species ranges is biased by the uneven attention paid to different organisms by researchers. Vertebrates and large organisms have been all-time favourites, in contrast to small and microscopic living forms that have remained neglected (Appeltans et al., 2012; Lamshead & Boucher, 2003; Vitorino & Bessa, 2018). Indeed, many aspects of the biology of minute organisms remain unknown (Fonseca et al., 2018) hampering our understanding of their spatial distribution patterns and the mechanisms that drive them (Fontaneto et al., 2012; Marrone et al., 2022; Martínez et al., 2019). This lack of knowledge seriously impacts our global understanding of biodiversity patterns because microscopic animals are a numerically important component of biodiversity in many regions (Curini-Galletti et al., 2012; Jörger et al., 2021; Martínez et al., 2019; Willems et al., 2009), offer essential services in many aquatic ecosystems (Schratzberger & Ingels, 2018) and represent tools to test general eco-evolutionary hypotheses avoiding the confounding factor of the non-independence of the observations due to shared evolutionary history (Fonseca et al., 2018). Furthermore, the anatomical simplification of organ systems due to miniaturisation depicts microscopic animals as excellent model organisms for a wide range of morphological to evolutionary, ecological or physiological studies. Many investigations have suggested that microscopic organisms exhibit much broader distribution ranges than their macroscopic counterparts (Fontaneto, 2011; Hillebrand, 2004; Maraun et al., 2007). These broader distributions are easy to explain for those microorganisms that possess traits favouring long-distance dispersal, such as high abundances, dormancy stages, long-term viability and parthenogenesis (Curini-Galletti et al., 2012; Fontaneto, 2019). What is more striking, however, is the large-scale ranges recorded for species without evident mechanisms of long-distance dispersal (Giere, 2008). This so-called 'meiofauna paradox' has been attributed to our inability to identify the actual units of diversity across wide areas: Genetically distinct species may exhibit highly conserved morphologies, masking geographical patterns of diversity. Additionally, species identifications are challenging for the non-expert, populating the literature with misidentifications that are seldom examined or rectified. Those wrongly attributed species names might then linger in the literature for years, causing additional inaccuracies in forthcoming studies. Indeed, sampling bias, which hassles all biodiversity inventories (Barbosa et al., 2010; Boakes et al., 2010), becomes hefty for meiofauna due to the problems inherent to the study of minute animals (Fontaneto et al., 2012). Some microscopic animals, though, are more popular than others, as most conventional ecological studies on meiofauna have focused on hard-bodied taxa rather than on groups with flimsy,

soft bodies (Curini-Galletti et al., 2020). This has been well documented in the literature and attributed to the fact that soft-bodied taxa (e.g. flatworms, gastrotrichs etc.) must be studied *in vivo* in the field, whereas hard-bodied animals (e.g. nematodes, arthropods) can be fixed and studied months after their collection. What remains to be addressed is what traits bias our knowledge of the distribution patterns within hard- and soft-bodied taxa in meiofauna.

In this study, we investigate the potential sources of bias affecting our knowledge of the distribution of microscopic animals using marine halacarids in Europe as a model group. Halacaridae is a family of microscopic mites that have been historically neglected, despite being relatively easy to find crawling in a variety of substrates, such as macroalgae, marine phanerogams, sandy and gravelly particles, or even associated with larger animals (Bartsch, 2006; Giere, 2008). We know to date more than 1200 species, predominantly marine although brackish and freshwater representatives have also been described (Bartsch, 2008). Halacarids are also common in a wide range of subterranean and semi-subterranean aquatic habitats, such as caves or the hyporheic zone associated with rivers (Husmann & Teschner, 1970; Schwoerbel, 1961, 1986). Halacarid mite species possess phytophagous, omnivorous and/or carnivorous diets, the latter including predators, parasites and probably scavengers. More striking is that many halacarid species, yet small (150–1000 µm) and only capable to crawl, have been recorded in all geographical provinces, from polar to tropical regions, and all depths, from the intertidal to abyssal zones (Bartsch, 2006; Newell, 1967). The degree of ecological specialisation, though, varies depending on the species, ranging from ecologically specialised forms to opportunistic groups recorded across a wide range of substrates. This versatility, along with their low dispersal capability and conserved morphology, depicts halacarids as a suitable model for addressing general questions in ecology and evolution (Pepato et al., 2019). Nevertheless, halacarid mites are understudied despite their widespread presence and many species have yet to be described (Appeltans et al., 2012). This leads to a unique situation amongst meiofaunal groups: being easy to preserve as well as geographically and ecologically widespread hard-bodied organisms, many experts on meiofauna have come across halacarids. Yet, only a few researchers have ventured into their taxonomy and provided records at the species level.

Our main premise is that sampling effort will largely explain the distribution patterns of these halacarid species, masking the ecological factors driving their distribution. This premise leads to three hypotheses. (H1) The patterns of taxonomic richness of halacarids will be explained by the presence of marine biological stations. (H2) Regardless of these biases, we expect that differences in species composition between marine provinces will explain the overall differences in species replacement across European marine stations. Yet, given the uneven sampling effort (Rubio-López et al., 2022), the species known in stations with fewer published papers and records will represent a subsample of the species known in the stations that have been better sampled. (H3) Finally, larger and generalist species might be easier to find and capable to thrive in more localities, accumulating more records than smaller or more ecologically specialised species.

2 | MATERIALS AND METHODS

2.1 | Data set and rationale

We focused on Europe because its coasts along the northern Atlantic and adjacent basins are the best investigated areas in the world for Halacaridae, with ca. 230 recorded marine species (Bartsch, 2004b), with the first records dating back from the 18th century (Baster, 1758).

We used a data set previously assembled (Rubio-López et al., 2022) with all occurrences of marine halacarids recorded between 70°N–20°N and 50°W–50°E. We divided our studied area according to marine provinces and countries following the Marine Ecoregions of the World (Spalding et al., 2007) and the Biodiversity Information Standards of the Taxonomic Database Working Group (www.tdwg.com). Accordingly, records from Svalbard and the deep sea (<200m depth) were considered outside the study area (coastal and shelf areas). Also, records from Macaronesia (Açores-Canary Islands) were removed because of the scarcity of records and they lack an effective marine station. Here, we refer to record as the report of a species at a physical location at a certain time (Isaac & Pocock, 2015).

2.2 | Hypothesis tests and statistical models

All analyses were performed using the statistical software R 4.1.2 (R Core Team, 2021). Before running each model, we excluded multicollinear explanatory variables (Pearson's $r > 0.6$) checking for correlation values using the package *psych* v. 2.2.5 (Revelle, 2022). We also examined the potential remaining effect of multicollinearity, together with the fit of each of our models, using the package *performance* v. 0.9.1 (Lüdtke et al., 2021).

2.2.1 | Species richness of marine halacarids in Europe is largely explained by sampling effort (H1)

We explored whether the distribution of marine biological stations explains the overall patterns of species richness exhibited by halacarid mites along the European coastline (H1). This is because most of the studies on marine halacarids in Europe were performed in littoral areas by researchers associated with marine facilities (e.g. Bartsch, 1975, 1976a, 1976b, 1977, 1979, 1980 at Roscoff). Therefore, we expect a decay in species richness as the distance from marine biological stations increases (H1a) but also the number of records and papers associated with each marine station will explain the differences in species richness (H1b).

We first calculated the distance of each record to the nearest marine station using the function 'spDistsN1' of the package *sp* v. 1.4–6 (Pebesma & Bivand, 2005). We assembled the data set of marine biological stations using the Mars network (The European Network of Marine Stations, <https://www.marinstations.org/>)

and CIESM (The Mediterranean Science Commission, <https://www.ciesm.org/>), completed by adding a few stations not included in those networks but historically relevant in halacarid research (Appendix S1). The coordinates for each station were extracted from Google Maps. We then grouped all species records according to the logarithm of the distance to the nearest marine station and modelled the effect of this distance on species richness using generalised linear models (GLM). We logarithmically transformed the distances because we expect that smaller distances near each station will have a stronger influence. In other words, we assume that scientists prefer sampling nearby a marine station, but as their sampling localities move away from a marine station, the importance of the distance would decay rapidly. Given the uncertainty in assigning coordinates for some records (Rubio-López et al., 2022), we did not use the exact coordinates, which could be biased, but alleviated the uncertainties by grouping the records in six distance bins, following the Sturges approximation, which determines the optimal number of classes in a frequency distribution. Each bin has been included as a continuous integer variable in our analyses. Records that were more than 400km away from a marine station were disregarded. As countries and marine provinces are collinear, we included provinces in the models A and B which is biogeographically meaningful to account for differences in species richness of marine halacarids; and countries in the model which we included the number of papers as a confounding factor, to account for differences in sampling effort amongst countries, since some of them belong to more than one marine province (e.g. France, Spain). In addition, as a metric of sampling intensity, we incorporated the number of halacarid records and published articles for each marine biological station. We used a negative binomial GLM, with the package *MASS* v. 7.3–57 (Ripley et al., 2013), checking for overdispersion. The significance of each independent variable was summarised as a Type II ANOVA table, using the function 'Anova' in the package *car* v. 3.0.9 (Fox et al., 2013). We further checked the relative importance of each explanatory variable with the 'varImp' function in the package *caret* v.6.0–90 (Kuhn et al., 2020).

Then, we investigated the factors explaining differences in species richness (H1b), after grouping the number of species according to their nearest station. We included the number of halacarid records and published papers associated with each marine biological station as a measure of sampling intensity. This association was inferred by geographical proximity. Based on that, each record could be unambiguously assigned to a single marine station. We also included the area of influence of each marine station, defined by the distance to its furthest record. We used generalised least squares models (GLS) to account for the potential effect of various spatial autocorrelation structures. The latter were calculated with the 'gls' function in the package *nlme* v. 3.1–153 (Pinheiro et al., 2007). We further checked the relative importance of each explanatory variable with the 'varImp' function in the package *caret* v.6.0–90 (Kuhn et al., 2020). We mapped the distribution of marine biological stations, as well as their number of records and number of species using the package *ggplot2* v. 3.3.5 (Wickham, 2016).

2.2.2 | Differences in species composition depend on both marine provinces and sampling effort (H2)

Despite biases in sampling effort and distances from marine stations, we still expect to find differences in species composition (i.e. beta diversity) across marine provinces based on their historical and biogeographic contexts. We decomposed beta diversity into its two components: species replacement and differences in species richness (Podani & Schmera, 2011). The substitution of species in one site by different species in another site results in species replacement (Qian et al., 2005). Many mechanisms (e.g. colonisation, dispersal limitation) can produce species loss or gain, resulting in differences in richness between sites (Carvalho et al., 2012). Separating the contributions of species richness differences and replacement to the total beta diversity might reveal distinct underlying processes concealing similarity in composition amongst locations or habitats. For instance, although some halacarid species occur across the Mediterranean, the Black Sea and the Atlantic Ocean (Bartsch, 1989, 2004a, 2004b), others are restricted to one of those seas (André, 1939; Bartsch, 1983, 1998). Due to the differences in the distribution area and range amongst those species, we expect to find a strong effect of the marine province and country in which each station is located on the total beta taxonomic diversity and its replacement component (H2a), reflecting the substitution of species with increasing geographical distances. In contrast, we expect that the number of records and published papers will have higher relative importance in explaining richness differences (H2b), as more species should be known from thoroughly studied areas.

First, we calculated the overall differences in total beta diversity and its two components across marine provinces. Then, we calculated the relative importance of the countries, the marine provinces, the number of records and the number of papers in explaining differences in beta diversity across marine stations using permutational analysis of variance (PERMANOVA; Anderson, 2001). To prevent stations with inadequate sampling, only the stations with more than 10 records were selected. Beta diversity was calculated as the Jaccard similarity index using the 'beta' function in the package *BAT* v. 2.7.1 (Cardoso et al., 2021); PERMANOVA was computed with the 'adonis' function of the package *vegan* v. 2.5-7 (Oksanen et al., 2020). The relationships between the differences in species composition amongst the four investigated provinces were graphically shown using a dendrogram, calculated with the function 'hclust' in the *stats* package (R Core Team, 2021) (see next section).

2.2.3 | Species traits influence the distribution patterns of halacarid species (H3)

We expect that larger and generalist species accumulate more records because they are easier to spot and capable to thrive in more localities. Hence, we expect that the most recorded species are larger and/or more able to inhabit different substrates, whereas small species and/or those inhabiting specific substrates (i.e. specialists) have fewer records (H3a). We further expect that this relationship will

affect our knowledge of the co-occurrence of species across marine stations because, in each of them, large and generalist species are more likely to be found first, more often and together with more species (H3b).

We assembled a data set with the number of records, number of habitats (i.e. types of substrates), number of depth zonation categories, number of countries, number of stations, number of localities each species has been reported for, as well as the idiosomal length (in μm ; Appendix S2) and the geographical range (in km) of each species. We have retrieved the body length from the illustrations available in the literature to make measurements of morphological traits, as used in other studies (Gonzalez et al., 2018). We only considered adults of each species, because illustrations of immature specimens are less frequent, and because different life stages show different ecological preferences and dispersal abilities, even within the same species (Bartsch, 2006). For species lacking illustrations, we used the size of the genus as estimation. Measurements were made using the free software 'ImageJ', widely used by experts in different areas to measure functional traits (Birn-Jeffery et al., 2012; Cobb & Sellers, 2020; Fowler et al., 2009; Tinius & Patrick Russell, 2017). We defined the geographical range for a species as the geodesic distance between the furthest sampling localities where the species has been found. The geographical range was calculated using the 'geodist' function in the package *geodist* v. 0.0.7 (Padgham et al., 2021). After checking for collinearity, we tested the effect of body size and the number of habitats against the number of records and the geographical range, only for species recorded more than once (H3a). We select a Gaussian distribution for the geographical range, after confirming the normal distribution of the residuals; and a negative binomial distribution for the number of records because it represents count data with overdispersion. We further checked the relative importance of each explanatory variable with the 'varImp' function in the package *caret* v.6.0-90 (Kuhn et al., 2020).

To investigate the effect of those traits on the differences in species composition across marine provinces (H3b), we first calculated the Bray-Curtis dissimilarity index (BC) between species in each marine province using the 'vegdist' function on the package *vegan* v. 2.5-7 (Oksanen et al., 2020). To compute the Bray-Curtis index, we used the species occurrence matrix of each marine province, which is equivalent to the transposed species community matrix. We selected the Bray-Curtis dissimilarity index because we wanted to account for the number of times that each species was recorded in the same marine province. Then, we calculated the relative importance of body length, number of habitats, number of records and species geographical range over the Bray-Curtis dissimilarity matrix using PERMANOVA (Anderson, 2001) with the 'adonis' function of the package *vegan* v. 2.5-7 (Oksanen et al., 2020). We visualised the number of times each species has been recorded in each province using a heat map drawn with the 'heatmap' function on the package *ComplexHeatmap* v. 2.10 (Gu et al., 2016). The number of records of each species was log-transformed to ease the visualisation of the results with a colour scale. Then, we represented the co-occurrence pattern of each species across the entire data set using the 'complete

linkage' hierarchical clustering method calculated with the function 'hclust' in the *stats* package version 4.1.2 (R Core Team, 2021).

3 | RESULTS

According to the first hypothesis (H1a), we expected a decay in species richness as the distance from marine biological stations increases. Our results supported this assumption, as we found a negative relationship between species richness and the distance to the closest marine station (GLM, estimate = -0.1 ± 0.045 , $t = -2.3$, $p = .024$). However, distance showed no significant effect when the number of records and papers published by each marine station is included in the model as confounding factors (Table 1). Differences in species richness across marine stations (H1b) were explained by the number of records, the number of publications and the area of influence of stations (Table 2), accounting for spatial structure with a rational quadratic spatial correlation structure. The distribution of the 141 marine biological stations, as well as the number of records and number of species for each of them, is shown in Figure 1.

Our results also supported our second hypothesis since, regardless of sampling bias, differences between marine provinces explained the highest percentage of the variance of the total beta diversity matrix and its replacement component (H2a, PERMANOVA, Table 3). In addition, as expected, the number of records had higher relative importance in explaining the variation in the richness component of the beta diversity (H2b; Table 3). The highest differences in total beta diversity across marine provinces (H2) ranged between 0.53 (Lusitania vs Northern European Seas) and 0.8 (Black Sea vs. Northern European Sea; Table S1). The cluster calculated from this matrix recovered the Northern European Seas as the first province branching off, followed by Lusitania

and a subcluster including the Mediterranean and the Black Sea (Figure 2). The dendrogram for the species matrix and the support values can be seen in the Figure S1.

Generalist species of halacarids (i.e. species occurring in more different types of substrates) accumulated more records and exhibited broader geographical ranges (H3a; Table 4). Conversely, species' body size did not show any relation with the number of records, but a marginally significant relationship with their geographical range (Table 4). The number of habitats and records accumulated per species were the predictors with the highest relative importance in explaining the species occurrence matrix (H3b, Table 5). We found congruent results for each marine province separately (Tables S2 and S3).

4 | DISCUSSION

Our analysis confirmed our main premise: sampling effort largely explains the differences in distribution patterns of marine mites in Europe. Specifically, we found a small yet significant decay of species richness with the distance from each marine station, whereas sampling effort explained differences in species richness across them (H1). Although most differences in total beta diversity and its replacement component were explained by countries and marine provinces, the number of papers and records had higher relative importance for the differences in beta richness (H2). Finally, generalist species (occurring in more habitats) had broader geographical ranges and accumulated more records, affecting the species occurrence known for each marine station (H3). However, sampling biases could not alone explain the distribution of records in our data set since, despite the bias, we found an effect of geography and species traits over those patterns. In other words, marine Halacaridae are

TABLE 1 Effect of distance, number of records and number of papers on species richness (A), number of records (B) and number of papers (C) of marine halacarids in Europe (H1a) according to a type II ANOVA output of generalised linear models.

Model	Response variable	Explanatory variables	LR Chisq	df	<i>p</i> -values	VarImp
A	Species richness	Distance bin	0.301	1	.583	0.540
		Records	118.337	1	<.001	13.899
		Papers	22.201	1	<.001	5.112
		Province	3.942	3	.268	3.752
B	Records	Distance bin	1.563	1	.211	1.294
		Papers	97.910	1	<.001	10.208
		Province	19.347	3	<.001	7.006
C	Papers	Distance bin	2.079	1	.149	1.426
		Country	38.910	17	.002	28.275

Note: The marine provinces and countries are collinear, so only one of them is included in each model. Models A and B include marine provinces as an explanatory variable, which is biogeographically meaningful to account for differences in species richness of marine halacarids; model C includes country to account for differences in sampling effort amongst countries, since some of them belong to more than one marine province (e.g. France, Spain). *p* values for significant predictors are marked in bold.

Abbreviations: df, degrees of freedom, LR Chisq, likelihood ratio chi-square values, VarImp, relative importance of each variable.

Explanatory variables	Estimate	Standard error	t-values	p-values	VarImp
(Intercept)	13.619	0.822	16.574	<.0001	
Area of influence	-1.405	0.256	-5.487	<.001	0.283
Records	7.294	0.763	9.562	<.001	5.946
Papers	5.990	0.745	8.045	<.001	4.126

Note: The number of records and papers are included as a measure of sampling efforts of the halacarid diversity across marine biological stations. The area of influence of each marine station is defined by the distance to its furthest record. The relative importance of each variable has also been calculated (VarImp). *p* values for significant predictors are marked in bold.

TABLE 2 Effect of sampling efforts (i.e. number of records and number of papers) on species richness of marine halacarids in Europe (H1b) according to a type II ANOVA output of generalised least squares.

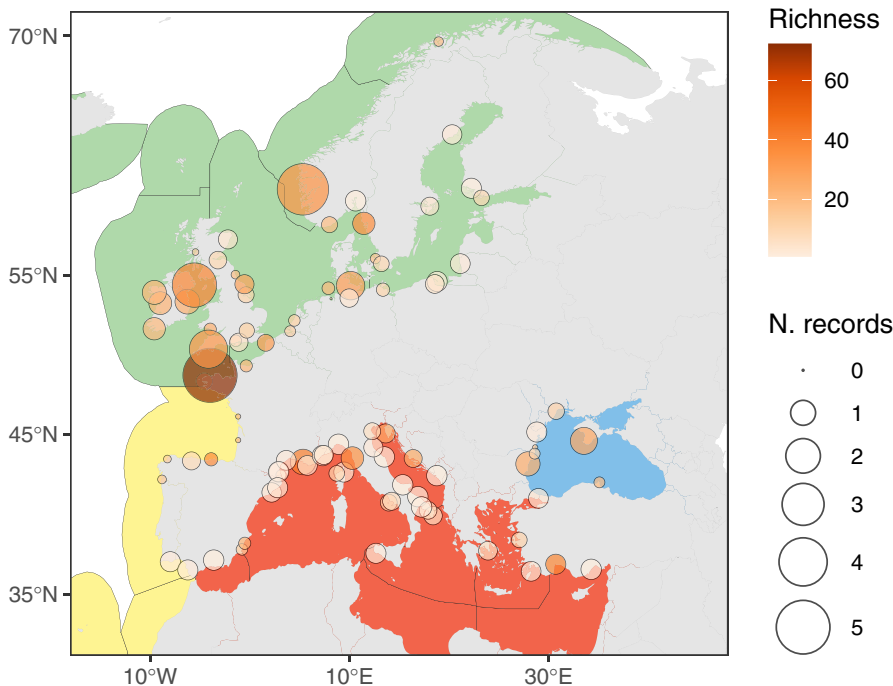


FIGURE 1 Distribution of marine halacarid mites (Halacaridae) recorded in Europe across the 141 marine biological stations studied here. Northern European Seas (NES) coloured in green; Lusitania (L) coloured in yellow; Mediterranean Sea (M) coloured in red; Black Sea (BS) coloured in blue. Point size varies according to the number of records, whereas point colour indicates the number of species.

not equally distributed across Europe, but their distribution patterns are affected by certain ecological and geographical processes.

4.1 | Impact of sampling effort on halacarid richness: The effect of marine stations

Our results confirmed that researchers focus on the surrounding areas of marine biological stations. This is not surprising, since biological stations play a pivotal role in addressing today's most critical biological issues in marine habitats (e.g. climate change, biodiversity loss, biological invasions) and provide a global network for long-term environmental monitoring and research, education and scientific dissemination (Struminger et al., 2018; Wilson, 1982; Wyman et al., 2009). Furthermore, marine stations also provide key resources for research, such as vessels and wet lab facilities, which are necessary for sample collection and species documentation. Consequently, biological stations might also pitch our knowledge on biodiversity leaving some areas unevaluated and potentially

introduce non-random geographical biases. Nevertheless, our results indicate only a moderate decline in species richness with increasing distance from research stations, but this result might be affected by two features of our data set. First, we enforced all records to a marine biological station, even when this might not always be the case. Thus, the enforcement might have inflated the number of records associated with certain isolated stations (e.g. Bergen as the only station in Western Norway), affecting the relationship between distance and species richness associated with the furthest bins. The second feature is related to the geographical uncertainty associated with those records that are not georeferenced (Marcer et al., 2022). In those cases, we inferred the coordinates after the centroid of the geographical information provided, which is further from the actual sampling site. We prefer this more conservative approach, though, rather than assuming that collections were always performed next to the station because we did not want to favour our working hypothesis a priori, but rather evaluate to what extent the hypothesis is sustained despite the issues implicit to the available data. Additionally, the variation in species richness was explained by

TABLE 3 Effect of sampling effort on species composition of marine halacarids (H2b).

Model	Variables	df	SS	R ²	F-values
β_{total}	Province	3	5.275	0.130	4.863
	Country	21	9.985	0.246	1.315
	Records	1	0.870	0.021	2.405
	Papers	1	0.566	0.014	1.567
	Residuals	66	23.863	0.588	
	Total	92	40.558	1.000	
β_{richness}	Province	3	1.498	0.087	4.235
	Country	21	3.945	0.230	1.593
	Records	1	2.366	0.138	20.068
	Papers	1	1.562	0.091	13.244
	Residuals	66	7.782	0.454	
	Total	92	17.153	1.000	
$\beta_{\text{replacement}}$	Province	3	1.773	0.161	4.712
	Country	21	3.064	0.279	1.163
	Records	1	-1.255	-0.114	-10.007
	Papers	1	0.862	-0.078	-6.875
	Residuals	66	8.279	0.753	
	Total	92	10.999	1.000	

Note: Results are reported from permutational multivariate analyses of variance (PERMANOVA) for the beta diversity (β_{total} , $\beta_{\text{replacement}}$, β_{richness}) across marine provinces. Explanatory variables include marine provinces, countries, number of records and number of papers.

Abbreviations: df, degrees of freedom; R² = coefficient of determination; SS, sum of squares.

the number of records and published articles associated with each marine station, indicating that our knowledge of diversity patterns is influenced by the number of researchers working in a given area (Fontaneto et al., 2012; Rubio-López et al., 2022). Sampling intensity does not drive actual species richness, but it does condition our knowledge of species richness and distribution patterns, acting as a confounding factor when patterns of diversity are to be inferred (Fontaneto et al., 2012).

4.2 | Occurrence of halacarid mites across different marine provinces: Ecology overrules sampling biases, to a certain extent

Macroecological variables explain differences in species composition of marine halacarids across the European seas, despite sampling bias. Accordingly, geography (i.e. marine province) accounts for most of the variance of the beta diversity matrix in our PERMANOVA analyses, particularly regarding differences in total beta diversity and its replacement component. We interpret this variation as the result of the interactions between ecological processes (e.g. historical constraints, environmental filtering, biological interactions and dispersal) with the limitations on our knowledge imposed by the unevenness in sampling effort. Hence, biological processes drive halacarid species composition, while sampling bias serves as a secondary filter, skewing our knowledge on the distribution of Halacaridae.

Looking at the differences in species composition across marine provinces (Figure 2), the Northern European Seas and Mediterranean provinces exhibit more species restricted to one of those geographical units in our data set (39 and 35, respectively). In contrast, most Lusitanian species are shared with either the Northern European Seas (e.g. *Halacarellus balticus*, *Thalassarachna basteri* and *Isobactrus setosus*) or with the Mediterranean (e.g. *Simognathus minutus*, *Copidognathus humerosus* and *Halacarus actenos*), suggesting that the species composition of Lusitania might be a subsample of the species present in the other two marine provinces. These results are surely affected by the distribution of data available for this province, with most of the records concentrated nearby the European continental coasts of Lusitania. In contrast, very little is known from the African coast or the Macaronesian archipelagos, which might potentially harbour very different species, and contribute to the number of endemic species in Lusitania (Freitas et al., 2019). The Black Sea has 13 endemic species, such as *Actacarus ponticus*, *Copidognathides ampliatus*, *Copidognathus mucronatus* and *Isobactrus peregrinus*.

4.3 | Generalist species accumulate more records: does this reflect the actual distribution patterns of mites or the biases produced by the sampling behaviour of researchers?

According to the results of our PERMANOVA, generalist species with a larger number of records are more likely to be found together,

Biogeography Habitat

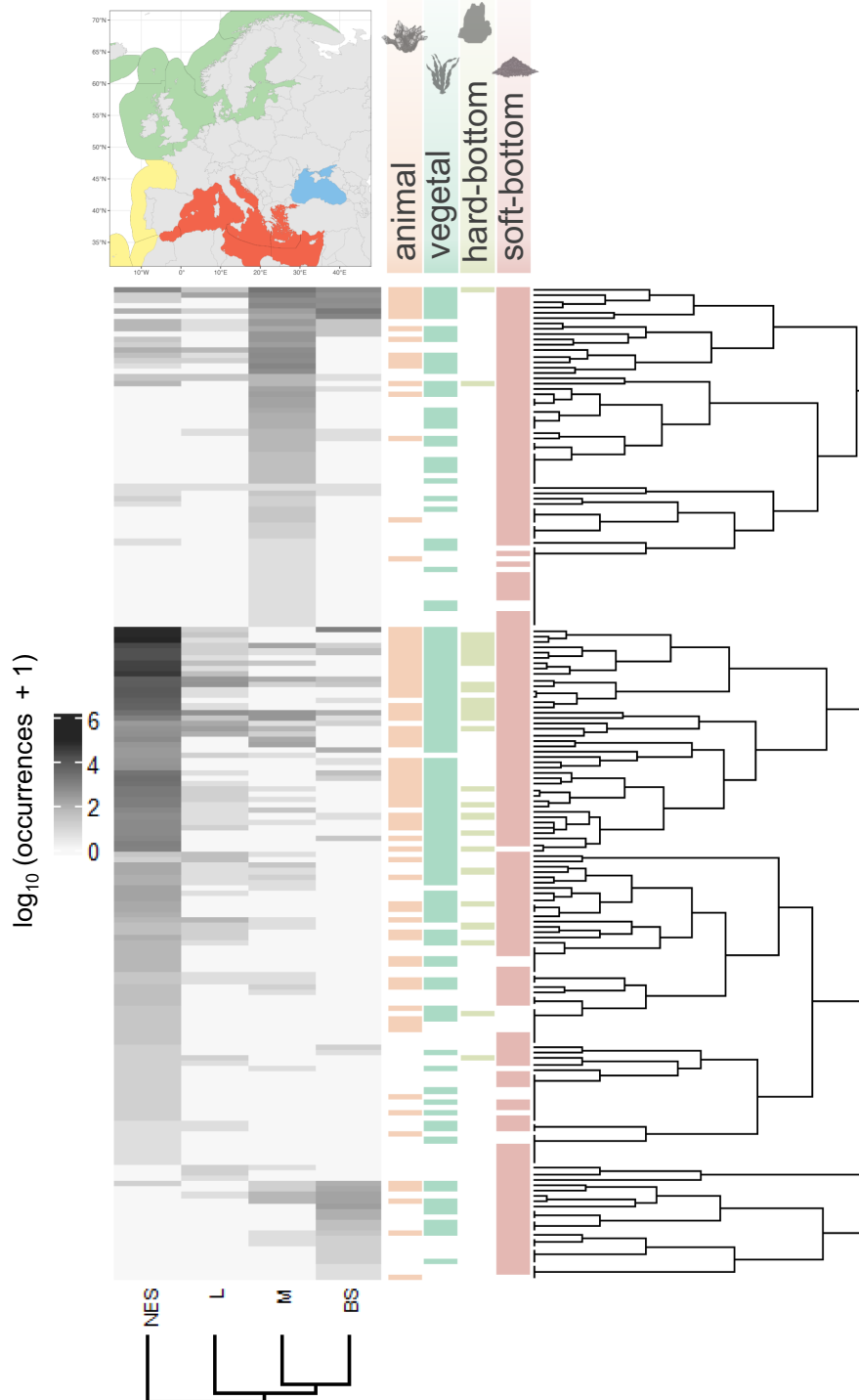


FIGURE 2 Heatmap showing the frequency at which each species has been recorded in each marine province. Each row corresponds to a species and each column to a marine province (Northern European Seas, Lusitania, Mediterranean and Black Sea respectively). We built the hierarchical clustering dendrogram on the right side using the corresponding values of the Bray–Curtis dissimilarity index obtained for each species. The distance has been calculated using the species occurrence matrix of each marine province, which is equivalent to the transposed species community matrix. The colour gradient from white to black indicates the number of occurrences (from low-white to high-dark). Each type of substrate is coloured: animal origin in watusi (orange), hard bottom in frost (light green), soft bottom in cavern pink and vegetal origin in jet stream (sky blue) respectively. Abbreviations as in [Figure 1](#).

independently of their body size. This indicates that species traits might shape the occurrence of species and ultimately the composition of mite communities across marine provinces.

However, these traits might also affect the detectability of different species of mites (Brown, 1984), which could confound our interpretations. Detectability refers to the probability of finding at least one individual of a given species in a sampling event, assuming that the species is present in the sampled region. We found a marginally

significant negative relationship between geographic range size and body size, but this is masked by sampling bias and cannot be trusted. Moreover, species' body size showed no correlation with the number of records in our analyses.

Conversely, generalist species collectively exhibit wider geographical ranges in our data set, being recorded more often than specialised taxa (i.e. occurring in fewer habitats). This may be because abundant species are more likely to appear in a given sample,

TABLE 4 Effect of species traits (number of habitats, body length) on the distribution of marine halacarids (H3a) according to a type II ANOVA output of generalised linear models.

Models	Variables	Estimate	Standard error	t-values	p-values	VarImp
Geographical range	(Intercept)	1656.228	299.722	5.526	<.001	
	Habitats	236.030	96.317	2.451	.015	2.451
	Length	-1.541	0.684	-2.252	.026	2.252
	Records	17.209	3.644	4.723	<.001	4.723
Number of records	(Intercept)	-0.326	0.184	-1.772	.076	
	Habitats	0.616	0.054	11.331	<.001	11.331
	Length	0.0003	0.0004	0.861	.390	0.861
	Max. distance	0.0001	0.0001	12.645	<.001	12.645

Note: Summary of the generalised linear models for the geographical range and the number of records, including the number of habitats as a proxy of the degree of generalism and the idiosomal length as a proxy of body size. *p* values for significant predictors are marked in bold. The relative importance of each variable has also been calculated (VarImp).

TABLE 5 Effect of species traits on our knowledge of the occurrence of species across marine provinces (H3b).

Variables	df	SS	R ²	F-values
Length	1	0.400	0.007	1.553
Habitats	1	7.112	0.118	27.611
Max. distance	1	1.198	0.033	7.689
Records	1	3.731	0.062	14.485
Residuals	182	46.879	0.780	
Total	186	60.103	1.000	

Note: Permutational multivariate analysis of variance (PERMANOVA) for the variables explaining differences in the distribution of marine mites. We used the following variables: idiosomal length as a proxy of the body size (length), number of habitats, number of records and the geographical range (max. distance) of each species. Abbreviations: df, degrees of freedom; R², coefficient of determination; SS, sum of squares.

increasing its apparent distribution range (Fridley et al., 2007; Gaston et al., 1997), and because scientists prefer to sample habitats where their target organisms are more common, inflating the rarity (sensu Brown, 1984) of specialised species.

In that regard, as researchers interested in biodiversity, we must be aware of two different cognitive biases that might affect the way we collect our data. One of these biases can be identified with the 'streetlight effect' or 'principle of the drunkard's search' (Kaplan, 1964) and it is represented in our analyses by the concentration of research on the surrounding areas of marine biological stations. Researchers presumably resort to sites close to marine stations because they are easily available and convenient. Interestingly, the fact that a given geographical area has been already well sampled might favour future researchers to continue sampling in it, if, for example, they are interested in collecting a particular species that was originally described in there for phylogenetics or physiological studies or to redescribe species with old, inaccurate descriptions or look for new type material if the original has been damaged or lost (Phillips et al., 2009; Reddy & Dávalos, 2003). This might reinforce

through time that areas known to have more species tend to attract more observers.

The second cognitive bias might be identified as the confirmation bias, inherent to the fact that many specialised taxonomists tend to sample more intensely those habitats where their target organisms are known to be more common. In that regard, scientists interested in Halacaridae might focus on vegetated areas because we know halacarids are abundant there (García-Gómez et al., 2022). Thus, habitats with high species richness might also comprise a disproportionately large amount of sampling effort. Similarly, specialists might prefer to use coarse sampling techniques in the field, such as dredging, because these often yield higher diversity by collecting across larger areas and by mixing substrates, rather than more precise methods that might yield fewer individuals but better ecological data (García-Gómez et al., 2022). Indeed, collecting 'a little bit here, a little bit there' is a common practice amongst meiobenthologists seeking to maximise sampling efficiency when producing diversity inventories (Curini-Galletti, pers. comm.).

Whereas these two biases might affect our knowledge of biodiversity across the tree of life, they might have a more accentuated impact on groups that require a higher specialised taxonomic training and equipment for their identification, such as meiofauna. In other words, many people, including researchers but also non-professional enthusiasts, can identify fish, birds, dragonflies or molluscs. This has been already exploited, for example, in several citizen science projects that have covered variable geographical scales, generating high-quality data and reliable scientific outcomes (Bartumeus et al., 2018; Južnič-Zonta et al., 2022; Wiggins & Crowston, 2012). In contrast, only a handful of highly specialised taxonomists can identify copepods, kinorhynchs or marine mites; not to mention other groups, such as free-living platyhelminths and acoels, for which very few taxonomists are currently active. Since some of those small groups of organisms are amongst the most important components of diversity in certain ecosystems, this problem must be taken seriously if we want to achieve the goal of having a complete picture of the Earth's biodiversity in the future.

5 | CONCLUSIONS

The variation in species richness was largely explained by uneven sampling effort, as indicated by the concentration of species known in the surrounding areas to marine biological stations and by a strong effect of the number of papers and records. However, our findings also supported that the species composition of marine halacarids varies across marine provinces, suggesting that ecological and historical factors explain the distribution of mites across large geographical scales. Nonetheless, sampling effort also explained the richness differences that emerge from the species sorting process. Finally, in our data set, generalist species have larger geographical ranges and are documented more frequently than ecologically specialised taxa. Our knowledge of the distribution patterns of marine mites largely depends on the patterns of sampling effort, which might conversely reflect the behaviour of the scientists studying these animals.

ACKNOWLEDGEMENTS

The MS has been written in the IRSA-CNR in Pallanza, supported by an Erasmus+ Mobility grant to IRL

CONFLICT OF INTEREST STATEMENT

All authors declare that they have no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/ddi.13679>.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the supplementary Information in Dryad (<https://doi.org/10.5061/dryad.d7wm37q53>). The data set is available to download in a previous article (Rubio-López et al., 2022) and can be accessed in <https://doi.org/10.21426/B637257011>.

REFERENCES

- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1), 32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>
- André, M. (1939). Description du *Coloboceras longiusculus* Trouessart (Halacarien). *Bulletin du Muséum National d'Histoire Naturelle*, 11, 402–404.
- Appeltans, W., Ah Yong, S. T., Anderson, G., Angel, M. V., Artois, T., Bailly, N., Bamber, R., Barber, A., Bartsch, I., Berta, A., Błażewicz-Paszkowycz, M., Bock, P., Boxshall, G., Boyko, C. B., Brandão, S. N., Bray, R. A., Bruce, N. L., Cairns, S. D., Chan, T.-Y., ... Costello, M. J. (2012). The magnitude of global marine species diversity. *Current Biology*, 22(23), 2189–2202. <https://doi.org/10.1016/j.cub.2012.09.036>
- Barbosa, A. M., Fontaneto, D., Marini, L., & Pautasso, M. (2010). Is the human population a large-scale indicator of the species richness of ground beetles? Carabidae, people and sampling effort. *Animal Conservation*, 13(5), 432–441. <https://doi.org/10.1111/j.1469-1795.2010.00363.x>
- Bartsch, I. (1975). A contribution to Rhombognathinen fauna (Halacaridae, Acari) of the Brittany coast. *Acarologia*, 17(1), 53–80.
- Bartsch, I. (1976a). Zur Systematik und Verbreitung der *Halacarellus*-Arten (Halacaridae, Acari) an der Bretagne-Küste. *Entomologische Mitteilungen Aus Dem Zoologischen Museum Hamburg*, 5(94), 97–109.
- Bartsch, I. (1976b). Contribution to Halacaridae fauna (Halacaridae, Acari) of the Brittany coast. Description of five species of the sand Lucke system. *Acarologia*, 17(4), 652–667.
- Bartsch, I. (1977). New species of *Lohmannella* (Halacaridae, Acari) from the Roscoff area, Brittany. *Cahiers de Biologie Marine*, 18, 141–153.
- Bartsch, I. (1979). Further studies on the *Copidognathus* fauna from the coast of Brittany. *Acarologia*, 20(2), 217–234.
- Bartsch, I. (1980). Halacaridae (Acari) in the bay of Morlaix (Brittany). *Acarologia*, 21(1), 34–45.
- Bartsch, I. (1983). *Copidognathus richardi* (Trouessart) (Halacaridae, Acari) a redescription. *Acarologia*, 24(3), 285–288.
- Bartsch, I. (1989). Marine mites (Halacaridae, Acari): A geographical and ecological survey. *Hydrobiologia*, 178(1), 21–42.
- Bartsch, I. (1998). First record of *Isobactrus* from the Black Sea (Acari, Halacaridae, Rhombognathinae). *Entomologische Mitteilungen Aus Dem Zoologischen Museum Hamburg*, 12, 269–276.
- Bartsch, I. (2004a). The Black Sea halacarid fauna (Halacaridae, Acari): Faunal comparison with the Mediterranean, eastern North Atlantic, North Sea, and Baltic and reflection on its origin. *Zoosystematics and Evolution*, 80(2), 143–158. <https://doi.org/10.1002/mmzn.20040800202>
- Bartsch, I. (2004b). Geographical and ecological distribution of marine halacarid genera and species (Acari: Halacaridae). *Experimental & Applied Acarology*, 34(1), 37–58. <https://doi.org/10.1023/B:APPA.0000044438.32992.35>
- Bartsch, I. (2006). Halacaridae (Acari): A guide to marine genera. *Organisms Diversity & Evolution*, 6(2), 125. <https://doi.org/10.1016/j.ode.2005.07.001>
- Bartsch, I. (2008). Global diversity of halacarid mites (Halacaridae, Acari: Arachnida) in freshwater. *Hydrobiologia*, 595(1), 317–322. <https://doi.org/10.1007/s10750-007-9026-0>
- Bartumeus, F., Oltra, A., & Palmer, J. R. B. (2018). Citizen science: A gateway for innovation in disease-carrying mosquito management? *Trends in Parasitology*, 34(9), 727–729. <https://doi.org/10.1016/j.pt.2018.04.010>
- Baster, J. (1758). Observations de Corallinis, iisque infidentibus polyphis, aliisque animalculis marinis. *Philosophical Transaction of the Royal Society of London*, 50, 258–279.
- Bik, H. M., Thomas, W. K., Lunt, D. H., & Lamshead, P. J. D. (2010). Low endemism, continued deep-shallow interchanges, and evidence for cosmopolitan distributions in free-living marine nematodes (order Enoplida). *BMC Evolutionary Biology*, 10(1), 389. <https://doi.org/10.1186/1471-2148-10-389>
- Birn-Jeffery, A. V., Miller, C. E., Naish, D., Rayfield, E. J., & Hone, D. W. E. (2012). Pedal claw curvature in birds, lizards and Mesozoic dinosaurs – Complicated categories and compensating for mass-specific and phylogenetic control. *PLoS One*, 7(12), e50555. <https://doi.org/10.1371/journal.pone.0050555>
- Boakes, E. H., McGowan, P. J. K., Fuller, R. A., Chang-qing, D., Clark, N. E., O'Connor, K., & Mace, G. M. (2010). Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLoS Biology*, 8(6), e1000385. <https://doi.org/10.1371/journal.pbio.1000385>
- Brown, J. H. (1984). On the relationship between abundance and distribution of species. *The American Naturalist*, 124(2), 255–279. <https://doi.org/10.1086/284267>
- Cardoso, P., Mammola, S., Rigal, F., & Carvalho, J. C. (2021). BAT: Biodiversity Assessment Tools (R package version 2.4. 2) [R]. <https://CRAN.R-project.org/package=BAT>
- Carvalho, J. C., Cardoso, P., & Gomes, P. (2012). Determining the relative roles of species replacement and species richness differences in generating beta-diversity patterns. *Global Ecology and Biogeography*, 21(7), 760–771. <https://doi.org/10.1111/j.1466-8238.2011.00694.x>

- Cobb, S. E., & Sellers, W. I. (2020). Inferring lifestyle for Aves and Theropoda: A model based on curvatures of extant avian ungual bones. *PLoS One*, 15(2), e0211173. <https://doi.org/10.1371/journal.pone.0211173>
- Curini-Galletti, M., Artois, T., Delogu, V., Smet, W. H. D., Fontaneto, D., Jondelius, U., Leasi, F., Martínez, A., Meyer-Wachsmuth, I., Nilsson, K. S., Tongiorgi, P., Worsaae, K., & Todaro, M. A. (2012). Patterns of diversity in soft-bodied Meiofauna: Dispersal ability and body size matter. *PLoS One*, 7(3), e33801. <https://doi.org/10.1371/journal.pone.0033801>
- Curini-Galletti, M., Artois, T., Di Domenico, M., Fontaneto, D., Jondelius, U., Jörger, K. M., Leasi, F., Martínez, A., Norenburg, J. L., Sterrer, W., & Todaro, M. A. (2020). Contribution of soft-bodied meiofaunal taxa to Italian marine biodiversity. *The European Zoological Journal*, 87(1), 369–384. <https://doi.org/10.1080/24750263.2020.1786607>
- Danovaro, R., & Gambi, C. (2022). Cosmopolitanism, rareness and endemism in deep-sea marine nematodes. *The European Zoological Journal*, 89(1), 653–665. <https://doi.org/10.1080/24750263.2022.2040621>
- Foissner, W. (2006). Biogeography and dispersal of micro-organisms: A review emphasizing Protists. *Acta Protozoologica*, 45(2), 111–136.
- Fonseca, G., Fontaneto, D., & Di Domenico, M. (2018). Addressing biodiversity shortfalls in meiofauna. *Journal of Experimental Marine Biology and Ecology*, 502, 26–38. <https://doi.org/10.1016/j.jembe.2017.05.007>
- Fontaneto, D. (2011). *Biogeography of microscopic organisms*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511974878>
- Fontaneto, D. (2019). Long-distance passive dispersal in microscopic aquatic animals. *Movement Ecology*, 7(1), 10. <https://doi.org/10.1186/s40462-019-0155-7>
- Fontaneto, D., Barbosa, A. M., Segers, H., & Pautasso, M. (2012). The 'rotiferologist' effect and other global correlates of species richness in monogonont rotifers. *Ecography*, 35(2), 174–182. <https://doi.org/10.1111/j.1600-0587.2011.06850.x>
- Fontaneto, D., Ficetola, G. F., Ambrosini, R., & Ricci, C. (2006). Patterns of diversity in microscopic animals: Are they comparable to those in protists or in larger animals? *Global Ecology and Biogeography*, 15(2), 153–162. <https://doi.org/10.1111/j.1466-822X.2006.00193.x>
- Fowler, D. W., Freedman, E. A., & Scannella, J. B. (2009). Predatory functional morphology in raptors: Interdigital variation in talon size is related to prey restraint and immobilisation technique. *PLoS One*, 4(11), e7999. <https://doi.org/10.1371/journal.pone.0007999>
- Fox, J., Friendly, M., & Weisberg, S. (2013). Hypothesis tests for multivariate linear models using the car package. *The R Journal*, 5(1), 39–52.
- Freitas, R., Romeiras, M., Silva, L., Cordeiro, R., Madeira, P., González, J. A., Wirtz, P., Falcón, J. M., Brito, A., Floeter, S. R., Afonso, P., Porteiro, F., Viera-Rodríguez, M. A., Neto, A. I., Haroun, R., Farminhão, J. N. M., Rebelo, A. C., Baptista, L., Melo, C. S., ... Ávila, S. P. (2019). Restructuring of the 'Macaronesia' biogeographic unit: A marine multi-taxon biogeographical approach. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-51786-6>
- Fridley, J. D., Vandermast, D. B., Kuppinger, D. M., Manthey, M., & Peet, R. K. (2007). Co-occurrence based assessment of habitat generalists and specialists: A new approach for the measurement of niche width. *Journal of Ecology*, 95(4), 707–722. <https://doi.org/10.1111/j.1365-2745.2007.01236.x>
- García-Gómez, G., García-Herrero, Á., Sánchez, N., Pardos, F., Izquierdo-Muñoz, A., Fontaneto, D., & Martínez, A. (2022). Meiofauna is an important, yet often overlooked, component of biodiversity in the ecosystem formed by *Posidonia oceanica*. *Invertebrate Biology*, 141(2), e12377. <https://doi.org/10.1111/ivb.12377>
- Garrick, R. C., Sands, C. J., Rowell, D. M., Tait, N. N., Greenslade, P., & Sunnucks, P. (2004). Phylogeography recapitulates topography: Very fine-scale local endemism of a saproxylic 'giant' springtail at Tallaganda in the Great Dividing Range of south-East Australia. *Molecular Ecology*, 13(11), 3329–3344. <https://doi.org/10.1111/j.1365-294X.2004.02340.x>
- Gaston, K. J. (2000). Global patterns in biodiversity. *Nature*, 405(6783) Article 6783, 220–227. <https://doi.org/10.1038/35012228>
- Gaston, K. J. (2009). Geographic range limits of species. *Proceedings of the Royal Society B: Biological Sciences*, 276(1661), 1391–1393. <https://doi.org/10.1098/rspb.2009.0100>
- Gaston, K. J., Blackburn, T. M., & Lawton, J. H. (1997). Interspecific abundance-range size relationships: An appraisal of mechanisms. *Journal of Animal Ecology*, 66(4), 579–601. <https://doi.org/10.2307/5951>
- Giere, O. (2008). *Meiobenthology: The microscopic motile Fauna of aquatic sediments*. Springer Science & Business Media. <https://link.springer.com/book/10.1007/978-3-540-68661-3>
- Gonzalez, B. C., Worsaae, K., Fontaneto, D., & Martínez, A. (2018). *Anophthalmia* and elongation of body appendages in cave scale worms (Annelida: Aphroditiformia). *Zoologica Scripta*, 47(1), 106–121. <https://doi.org/10.1111/zsc.12258>
- Gotelli, N. J., Anderson, M. J., Arita, H. T., Chao, A., Colwell, R. K., Connolly, S. R., Currie, D. J., Dunn, R. R., Graves, G. R., Green, J. L., Grytnes, J.-A., Jiang, Y.-H., Jetz, W., Kathleen Lyons, S., McCain, C. M., Magurran, A. E., Rahbek, C., Rangel, T. F. L. V. B., Soberón, J., ... Willig, M. R. (2009). Patterns and causes of species richness: A general simulation model for macroecology. *Ecology Letters*, 12(9), 873–886. <https://doi.org/10.1111/j.1461-0248.2009.01353.x>
- Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18), 2847–2849. <https://doi.org/10.1093/bioinformatics/btw313>
- Hawkins, B. A., Field, R., Cornell, H. V., Currie, D. J., Guégan, J.-F., Kaufman, D. M., Kerr, J. T., Mittelbach, G. G., Oberdorff, T., O'Brien, E. M., Porter, E. E., & Turner, J. R. G. (2003). Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, 84(12), 3105–3117. <https://doi.org/10.1890/03-8006>
- Hillebrand, H. (2004). On the generality of the latitudinal diversity gradient. *The American Naturalist*, 163(2), 192–211. <https://doi.org/10.1086/381004>
- Husmann, S., & Teschner, D. (1970). Ökologie, Morphologie und Verbreitungsgeschichte subterranean Wassermilben (Limnolhalacaridae) aus Schweden. *Archiv für Hydrobiologie*, 67(2), 242–267.
- Isaac, N. J. B., & Pockock, M. J. O. (2015). Bias and information in biological records. *Biological Journal of the Linnean Society*, 115(3), 522–531. <https://doi.org/10.1111/bj.12532>
- Jörger, K., Alvaro, N., Andrade, L., Araujo, T., Aramayo, V., Artois, T., Ballentine, W., Bergmeier, F., Botelho, A., Buckenmeyer, A., Capucho, A., Cherenva, I., Curini-Galletti, M., Davison, A., Deng, W., Di Domenico, M., Ellison, C., Engelhardt, J., Fais, M., ... Ricardo Costa, A. C. (2021). Meiozoos 2019—exploring the marine meiofauna of the azores. *Açoreana, Suplemento* 11, 17–41.
- Južnič-Zonta, Ž., Sanpera-Calbet, I., Eritja, R., Palmer, J. R. B., Escobar, A., Garriga, J., Oltra, A., Richter-Boix, A., Schaffner, F., Torre, A., della, Miranda, M. Á., Koopmans, M., Barzon, L., Ferre, F. B., Alarcón-Elbal, P. M., González, M. A., Puig, M. A., Bakran-Lebl, K., Balatsos, G., ... Community, M. A. (2022). Mosquito alert: Leveraging citizen science to create a GBIF mosquito occurrence dataset. *Gigabyte*, 2022, 1–11. <https://doi.org/10.46471/gigabyte.54>
- Kaplan, A. (1964). *The conduct of inquiry*. Transaction Publishers.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., & Kenkel, B. (2020). *Caret: Classification and regression training (R package version 6.0–86)*. Astrophysics Source Code Library, ascl-1505.003. <https://github.com/topepo/caret/>
- Lamshead, P. J. D., & Boucher, G. (2003). Marine nematode deep-sea biodiversity – Hyperdiverse or hype? *Journal of Biogeography*, 30(4), 475–485. <https://doi.org/10.1046/j.1365-2699.2003.00843.x>
- Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6(60), 3139. <https://doi.org/10.21105/joss.03139>

- MacArthur, R. H., & Wilson, E. O. (2016). *The theory of Island biogeography*. Princeton University Press. <https://doi.org/10.1515/9781400881376>
- Maraun, M., Schatz, H., & Scheu, S. (2007). Awesome or ordinary? Global diversity patterns of oribatid mites. *Ecography*, 30(2), 209–216. <https://doi.org/10.1111/j.0906-7590.2007.04994.x>
- Marcet, A., Chapman, A. D., Wiczorek, J. R., Xavier Picó, F., Uribe, F., Waller, J., & Ariño, A. H. (2022). Uncertainty matters: Ascertaining where specimens in natural history collections come from and its implications for predicting species distributions. *Ecography*, 2022(9), e06025. <https://doi.org/10.1111/ecog.06025>
- Marrone, F., Fontaneto, D., & Naselli-Flores, L. (2022). Cryptic diversity, niche displacement and our poor understanding of taxonomy and ecology of aquatic microorganisms. *Hydrobiologia*, 849, 1–16. <https://doi.org/10.1007/s10750-022-04904-x>
- Martínez, A., Di Domenico, M., Leasi, F., Curini-Galletti, M., Todaro, M. A., Zotto, M. D., Gobert, S., Artois, T., Norenburg, J., Jörgen, K. M., Núñez, J., Fontaneto, D., & Worsaae, K. (2019). Patterns of diversity and endemism of soft-bodied meiofauna in an oceanic island, Lanzarote. *Canary Islands. Marine Biodiversity*, 49(5), 2033–2055. <https://doi.org/10.1007/s12526-019-01007-0>
- Newell, I. M. (1967). Abyssal Halacaridae (Acari) from the Southeast Pacific. *Pacific Insects*, 9(4), 16.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2020). *Vegan: Community ecology package* (R package version 2.5-7) [R]. <https://CRAN.R-project.org/package=vegan>
- Padgham, M., Summer, M. D., & Karney, C. F. F. (2021). *Geodist: Fast. Dependency-Free Geodesic Distance Calculations* (R package version 0.0.7). <https://CRAN.R-project.org/package=geodist>
- Pebesma, E., & Bivand, R. S. (2005). Classes and methods for spatial data: The sp package. *R News*, 5(2), 9–13.
- Pepato, A. R., Vidigal, T. H. D. A., & Klimov, P. B. (2019). Evaluating the boundaries of marine biogeographic regions of the southwestern Atlantic using halacarid mites (Halacaridae), meiobenthic organisms with a low dispersal potential. *Ecology and Evolution*, 9(23), 13359–13374. <https://doi.org/10.1002/ece3.5791>
- Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, 19(1), 181–197. <https://doi.org/10.1890/07-2153.1>
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. C. (2007). *Nlme: Linear and Nonlinear Mixed Effects Models. R Package Version*, 3(57), 1–89. <https://CRAN.R-project.org/package=nlme>
- Podani, J., & Schmera, D. (2011). A new conceptual and methodological framework for exploring and explaining pattern in presence – Absence data. *Oikos*, 120(11), 1625–1638. <https://doi.org/10.1111/j.1600-0706.2011.19451.x>
- Qian, H., Ricklefs, R. E., & White, P. S. (2005). Beta diversity of angiosperms in temperate floras of eastern Asia and eastern North America. *Ecology Letters*, 8(1), 15–22. <https://doi.org/10.1111/j.1461-0248.2004.00682.x>
- Reddy, S., & Dávalos, L. M. (2003). Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography*, 30(11), 1719–1727. <https://doi.org/10.1046/j.1365-2699.2003.00946.x>
- Revelle, W. (2022). *Procedures for psychological, psychometric and personality research*. Psych (R package version 2.2.5) [R]. Northwestern University. <https://CRAN.R-project.org/package=psych>
- Ribera, I., Beutel, R. G., Balke, M., & Vogler, A. P. (2002). Discovery of Aspidytidae, a new family of aquatic Coleoptera. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 269(1507), 2351–2356. <https://doi.org/10.1098/rspb.2002.2157>
- Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). *Package 'mass' Cran r* 538, 113–120.
- Rubio-López, I., Pardos, F., Martínez, A., & García-Gómez, G. (2022). A dataset of European marine mites (Trombidiformes, Halacaridae). *Biogeographia - The Journal of Integrative Biogeography*, 37(2), 1–18. <https://doi.org/10.21426/B637257011>
- Schemske, D. W. (2002). *Ecological and evolutionary perspectives on the origins of tropical diversity* (pp. 163–173). Foundations of Tropical Forest Biology.
- Schratzberger, M., & Ingels, J. (2018). Meiofauna matters: The roles of meiofauna in benthic ecosystems. *Journal of Experimental Marine Biology and Ecology*, 502, 12–25. <https://doi.org/10.1016/j.jembe.2017.01.007>
- Schwoerbel, J. (1961). Subterrane Wassermilben (Acari: Hydrachnellae, Porohalacaridae und Stygothrombiidae), ihre Ökologie und Bedeutung für die Abgrenzung eines aquatischen Lebensraums zwischen Oberfläche und Grundwasser. *Archiv für Hydrobiologie, Supplement*, 25(4), 2–3.
- Schwoerbel, J. (1986). *Acari: Limnohalacaridae and Hydrovolziidae* (pp. 643–647). Stygofauna Mundi.
- Smith, H. G., & Wilkinson, D. M. (2007). Not all free-living microorganisms have cosmopolitan distributions – The case of *Nebela* (*Apodera*) vas certes (protozoa: Amoebozoa: Arcellinida). *Journal of Biogeography*, 34(10), 1822–1831. <https://doi.org/10.1111/j.1365-2699.2007.01733.x>
- Spalding, M. D., Fox, H. E., Allen, G. R., Davidson, N., Ferdaña, Z. A., Finlayson, M., Halpern, B. S., Jorge, M. A., Lombana, A., Lourie, S. A., Martin, K. D., McManus, E., Molnar, J., Recchia, C. A., & Robertson, J. (2007). Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. *Bioscience*, 57(7), 573–583. <https://doi.org/10.1641/B570707>
- Struminger, R., Zarestky, J., Short, R. A., & Lawing, A. M. (2018). A framework for informal STEM education outreach at Field stations. *Bioscience*, 68(12), 969–978. <https://doi.org/10.1093/biosci/biy108>
- Team, R. C. (2021). *Version 4.1.2. R: A language and environment for statistical computing* (4.1.2). R Foundation for Statistical Computing.
- Tinius, A., & Patrick Russell, A. (2017). Points on the curve: An analysis of methods for assessing the shape of vertebrate claws. *Journal of Morphology*, 278(2), 150–169. <https://doi.org/10.1002/jmor.20625>
- Vitorino, L. C., & Bessa, L. A. (2018). Microbial diversity: The gap between the estimated and the known. *Diversity*, 10(2), 46. <https://doi.org/10.3390/d10020046>
- Wickham, H. (2016). Data Analysis. In H. Wickham (Ed.), *Ggplot2: Elegant graphics for data analysis* (pp. 189–201). Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4_9
- Wiggins, A., & Crowston, K. (2012). *Goals and tasks: Two typologies of citizen science projects*. 2012 45th Hawaii international conference on system sciences 3426–3435. <https://doi.org/10.1109/HICSS.2012.295>
- Willems, W. R., Curini-Galletti, M., Ferrero, T. J., Fontaneto, D., Heiner, I., Huys, R., Ivanenko, V. N., Kristensen, R. M., Kånneby, T., MacNaughton, M. O., Martínez-Arbizu, P., Todaro, M. A., Sterrer, W., & Jondelius, U. (2009). Meiofauna of the Koster-area, results from a workshop at the Sven Lovén Centre for Marine Sciences (Tjärnö, Sweden). *Meiofauna Marina*, 17, 1–34.
- Willig, M. R., Kaufman, D. M., & Stevens, R. D. (2003). Latitudinal gradients of biodiversity: Pattern, process, scale, and synthesis. *Annual Review of Ecology, Evolution, and Systematics*, 34(1), 273–309. <https://doi.org/10.1146/annurev.ecolsys.34.012103.144032>
- Wilson, E. O. (1982). The importance of biological field stations. *Bioscience*, 32(5), 320. <https://doi.org/10.2307/1308847>
- Wyman, R. L., Wallensky, E., & Baine, M. (2009). The activities and importance of international field stations. *Bioscience*, 59(7), 584–592. <https://doi.org/10.1525/bio.2009.59.7.9>

BIOSKETCH

Iñigo Rubio-López is a predoc student who recently finished his master thesis at the Universidad Complutense de Madrid and a research stay at the National Research Council of Italy. He has a major interest in biogeography, evolution and ecology of microscopic animals, especially aquatic mites. Alejandro Martínez is a researcher at the Molecular Ecology Group (MEG) at the National Research Council of Italy (CNR-IRSA). His research focuses on the eco-evolutionary processes that affect microscopic animals, especially those living between sand grains, and animals from marine or freshwater flooded caves. Guillermo García-Gómez is pursuing a PhD at the University of Liverpool, with an interest in ecology and evolution, as well as systematics of marine mites.

Author contributions: IRL, GGG and AM conceived and designed the study; IRL and AM analysed the data with inputs from DF and GGG; IRL wrote the article with significant input from AM. FP supported the data collection and provided insightful comments on the text. All authors read and approved the final version of the manuscript.

SUPPORTING INFORMATION

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

How to cite this article: Rubio-López, I., Pardos, F., Fontaneto, D., Martínez, A., & García-Gómez, G. (2023). Biases and distribution patterns in hard-bodied microscopic animals (Acari: Halacaridae): Size does not matter, but generalism and sampling effort do. *Diversity and Distributions*, 00, 1–13. <https://doi.org/10.1111/ddi.13679>