Coral reefs provide cover for many important fish species that help maintain ecological balance (Sano et al. 1987, Sano 2000, Pratchett 2007). It is worth mentioning that biologically diverse communities ecologically emphasize functional redundancy, where species with the same function can support absent species or species undergoing recovery to not only maintain a properly functioning ecosystem but to maximize the biomass of species in that ecosystem (Lefcheck et al. 2021). There are many coral reef communities all over the world that contain many different species. For example, reef fish communities in the Caribbean region can contain anywhere from about 500-700 species, whereas, in the central pacific, they can range from 3000 species and over (Goldman & Talbot 1976, Sale 1980). Many different studies have compared species richness and abundance of reef fish within-habitat and between habitats on a continental to regional scale. However, a comparison of species abundance and richness between the 12 global realms has yet to be analyzed. As a result, this project seeks to compare the species richness across the different realms around the world. It is worth noting that species richness in a sample can greatly be influenced by sample size and sample effort of researchers based on certain factors within a realm. For instance, the sampling effort in the Arctic realm might be way less than the sampling effort in the tropics due to the sub-zero temperatures restricting divers. In order to make a fair comparison and assessment of the diversities across the multiple abundance counts within each of the 12 realms, rarefaction will be used to standardize the larger samples until they contain the same number of observed individuals or observations as the smallest sample. The objectives of this project are to: determine which of the 12 realms has the lowest and the highest number of individuals observed standardized to the same area. Determining the lowest and richest realms after standardizing per number of individuals, and determining the lowest and richest locality after standardizing according to sample coverage.

Methodology

The dataset for this project focuses on the systematic global assessment of reef fish communities by the Reef Life Survey program, which was collected by Graham J Edgar and Rick D Stuart-Smith in 2014. In the original dataset, the authors recorded the abundance of 2367 reef fish species across 11 realms. Other information such as the family the species came from, and the transects, sites, countries, and ecoregions where they were found were also recorded (15 columns and 134759 rows). This dataset was manipulated to contain only the 11 realms and accumulated abundances for the 2367 distinct reef fish species. This was to carry out our rarefaction and extrapolation for the abundance of species across the 11 realms. This analysis was done using the iNEXT package by Anne Chao (site) in the R software (Version 2022.12.0+353 (2022.12.0+353).

After manipulating the dataset, the integrated sample-sized-based rarefaction and extrapolation Hill numbers for each of the realms were calculated then integrated curves were constructed.

The coverage-based diversities of the 11 realms were then compared for species richness (q=0), Shannon diversity (q=1), and Simpson's diversity (q=2) according to the ‘‘base coverage’’ which is the lowest coverage for doubled reference sample sizes.

Results and Discussion

After comparing the 11 assemblages (realms) with Hill number order q = 0, 1, 2. The basic data information revealed that the lowest number of individuals observed among the 11 realms was 18 individuals in the Arctic realm. Whereas, the highest was 1,498,462 individuals from the Temperate Australasia region. The lowest richness observed among the 11 assemblages standardized to the same area was 5 species, which again came from the Artic realm. However, for the highest richness observed, the Central Indo-Pacific topped the realms with 1146 species.

After comparing sample-size-based sampling curves according to a base sample size, for each of the 11 realms, the integrated sample-size based rarefaction and extrapolation curves for Hill revealed that

The curve for species richness (q ¼ 0) increases steeply with sample size in both treatments, but the curves for Shannon and Simpson diversity (q¼1 and q¼ 2) level off beyond the reference sample, illustrating that higher order Hill numbers are increasingly dominated by the frequencies of the more common species and are, therefore, less sensitive to sampling effects. To compare diversities between the girdled and logged

For this rarefied sample, the Hill numbers of q ¼ 0, 1, 2 are estimated to be 31.71, 13.83, and 6.68, respectively. The proposed integrated sampling curve allows reliable comparisons for any sample size up to an abundance of 336. Across this range of abundance, Fig. 3b reveals that the logged treatment is more diverse for all but the smallest sample sizes for species richness (q ¼ 0) and Shannon diversity (q ¼ 1), although the confidence intervals overlap. In contrast, for Simpson diversity (q ¼ 2), the girdled treatment is more diverse, although again the two confidence intervals overlap.

Compare coverage-based sampling curves up to a ‘‘base coverage’’ (Fig

, we compare the coverage-based diversities of the two treatments for q ¼ 0 (left panel), q ¼ 1 (middle panel), and q¼2 (right panel) up to the coverage of 96%. This is our ‘‘base coverage’’ (the lowest coverage for doubled reference sample sizes or the maximum coverage for reference samples, whichever is larger)

the girdled treatment for any standardized sample coverage between 50% and 96%. For Shannon diversity (q ¼ 1), the logged treatment is more diverse, but the confidence bands overlap. For Simpson diversity (q ¼ 2), when coverage is less than 70%, both treatments have almost the same diversity, but when coverage is greater than 70%, the Simpson diversity for the girdled treatment is slightly higher. Comparing Figs. 3b and 5b, we see that the sample

The correlation between fish-species richness and coral generic richness (r = 0.65, p < 0.0001; electronic supplementary material, figure S1), which is consistent with reef-scale observations that fish richness increases with coral diversity( Giovanni Strona 2021)

We used this model to predict site-level fish species richness with and without corals. A fish community’s vulnerability to coral loss varied geographically, being highest in the central Pacific, intermediate in the western Indian, central Indo-Pacific, and tropical eastern Pacific, and moderate in the western Atlantic (figure 2a,c), ( Giovanni Strona 2021)

There is evidence to suggest that live coral cover is important for many fish species, such as obligate corallivores (Sano et al. 1987, Sano 2000, Pratchett 2007) or those intimately associated in their early life history (Jones et al. 2004). However, there is also evidence that topographic complexity is important for the provision of shelter (see Sano 2000). Topographic complexity is defined here as the sum of complexity afforded by living and dead coral (structural complexity) and complexity of the underlying reef matrix (substrate complexity). It is often unclear whether structural or substrate complexity is more important in determining fish communities, and we avoid this confusion by simply referring to topographic complexity

We explored several different model structures for describing fish diversity while accounting for major environmental and biogeographic factors that affect fish and coral diversity, namely mean surface water temperature; temperature annual range; salinity; pH; primary productivity; fraction of 1° × 1° cell area within 0–30 m; fraction of cell area intersecting reef habitat (according to http://data.unep-wcmc. org/datasets/1); isolation (total land area surrounding a reef locality within a radius of 5° latitude/longitude); marine biogeographic region of belonging [17]; absolute latitude. The environmental data (surface temperature, salinity, pH and total chlorophyll as a proxy for productivity) were obtained from [32]. Reef area and shallow habitat area helped account for expected species–area relationships that drive diversity. (Giovanni Strona)

Based on the above definition, our goal is to construct a diversity accumulation curve as a function of sample size (the number of individuals for abundance data or the number of sampling units for incidence data) or sample completeness.

We used the data from these two treatments to illustrate the construction of two types (sample-size- and coverage-based) of rarefaction and extrapolation curves of Hill numbers. The constructed sampling curves were then used to compare spider species diversities between the two treatments.

In comparing diversities among multiple assemblages, samples can be standardized by sample size or by sample completeness.

Coverage is defined as the total relative abundances of the observed species, or equivalently, the proportion of the total number of individuals in an assemblage that belong to species represented in the sample.

The estimated complement of coverage is not an estimate of the number of unseen species, but rather it estimates the proportion of the total individuals in the assemblage that belong to undetected species. For this reason, extremely rare, undetected species do not make a significant contribution to that proportion, even if there are many such species. This intuitively explains why the estimation of species richness in highly diverse assemblages is a statistically difficult issue, whereas sample coverage can be accurately estimated. Alroy (2010) and Jost (2010) independently proposed

The girdled observed species richness, Shannon diversity, and Simpson diversity (i.e., Hill numbers for q ¼ 0, 1, 2) for this reference sample size were, respectively, 26, 12.06, and 7.84 (solid points in Fig. 3a and b). The sample size for the logged treatment was 252, and the corresponding observed Hill numbers for q¼0, 1, 2 were 37, 14.42, and 6.76, respectively. Thus, judging from the unstandardized raw data (the reference samples), the logged treatment appears to have higher observed species richness and Shannon diversity, but lower Simpson diversity than the girdled treatment.

Step 1: Compare sample-size-based sampling curves

The curve for species richness (q=0) after standardizing per number of individuals (36) increases steeply with sample size in all assemblages except the Arctic, Temperate South America, Temperate South Africa and the Temperate Northern Pacific realms indicating that these realms had the lowest species richness of all the 11 realms whereas, Central Indo Pacific and Temperate Australasia dominated in terms of species richness. Furthermore, the curves for Shannon and Simpson diversity (q=1 and q= 2 respectively) leveled off just beyond the reference sample (36).

indicating that higher order Hill numbers are increasingly dominated by the frequencies of the more common species and are, therefore, less sensitive to sampling effects. To compare diversities between the girdled and logged

For this rarefied sample, the Hill numbers of q ¼ 0, 1, 2 are estimated to be 31.71, 13.83, and 6.68, respectively. The proposed integrated sampling curve allows reliable comparisons for any sample size up to an abundance of 336. Across this range of abundance, Fig. 3b reveals that the logged treatment is more diverse for all but the smallest sample sizes for species richness (q ¼ 0) and Shannon diversity (q ¼ 1), although the confidence intervals overlap. In contrast, for Simpson diversity (q ¼ 2), the girdled treatment is more diverse, although again the two confidence intervals overlap. Fig. 3b reveals that the logged treatment is more diverse for all but the smallest sample sizes for species richness (q ¼ 0) and Shannon diversity (q ¼ 1), although the confidence intervals overlap. In contrast, for Simpson diversity (q ¼ 2), the girdled treatment is more diverse, although again the two confidence intervals overlap.

Construct a sample completeness curve to link sample-size- and coverage-based sampling curves (Fig. 4).—Based on Eq. 12, the coverage for the girdled treatment is estimated as 93% for the reference sample of size 168 individuals, and the coverage for the logged treatment is 94% for the reference sample of 252 individuals. It is informative to examine how the sample

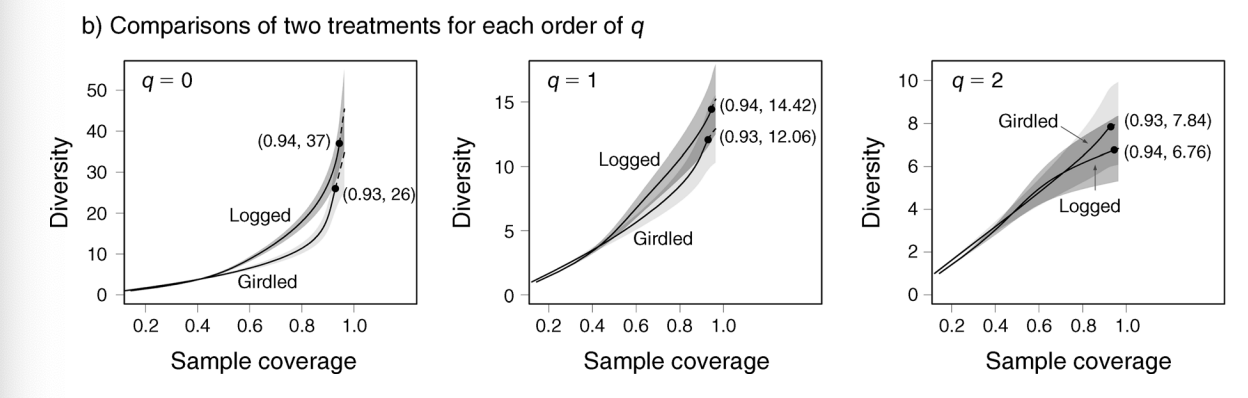
For any sample size less than 168, the curve shows that the sample completeness for the girdled treatment is estimated to be higher than that in logged treatment, although the

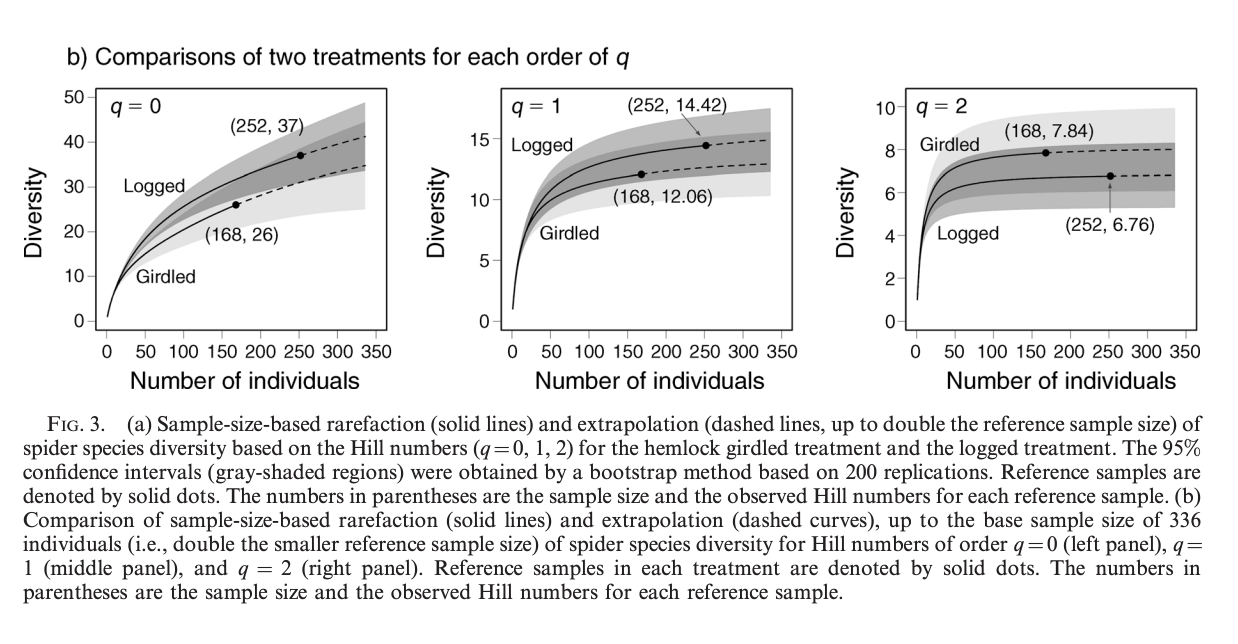
With such a curve, ecologists can objectively quantify the sample completeness for any incomplete abundance or incidence data sets. These curves help determine a sample size needed in a designing a survey. confidence intervals overlap. When sample size is larger than 168, the estimates of sample coveragesthe coverage for the girdled treatment is estimated as 93% for the reference sample of size 168 individuals, and the coverage for the logged treatment is 94% for the reference sample of 252 individuals. It is informative to examine how the sample completeness varies with sample size (see the formulas in the last row in Table 1). In Fig. 4, we plot the sample completeness curve as a function of sample size for each of the two treatments, up to double the reference sample size. For any sample size less than 168, the curve shows that the sample completeness for the girdled treatment is estimated to be higher than that in logged treatment, although the confidence intervals overlap. When sample size is larger than 168, the estimates of sample coverages

when sample size in the girdled treatment is doubled from 168 to 336 individuals, the sample coverage is increased from 93% to 96%. In the logged treatment, when sample size is doubled from 252 to 504 individuals, the coverage is increased from 94% to 97%. In Fig. 5a, we present, for each treatment, the corresponding coverage-based rarefaction and extrapolation curves with 95% confidence intervals for diversity of q ¼ 0, 1, 2 when the coverage is extrapolated to the value for a doubling of each reference sample size. we compare the coverage-based diversities of the two treatments for q ¼ 0 (left panel), q ¼ 1 (middle panel), and q¼2 (right panel) up to the coverage of 96%. This is our ‘‘base coverage’’ (the lowest coverage for doubled reference sample sizes or the maximum coverage for reference samples, whichever is larger). See Box 1 for suggestions on the choice of base coverage. Because the increase in coverage for the extrapolation is small, and the estimated diversity for q ¼ 1 and 2 hardly

change beyond the reference samples, the extrapolation parts in Fig. 5b are nearly invisible for these two orders of q. Since the two confidence bands do not intersect for species richness (q ¼ 0) if coverage exceeds 50% (Fig. 5b, left panel), species richness in the Central Indo Pacific realm is significantly higher than in any other realm for any standardized sample coverage up to 99%. For Shannon diversity (q=1), the the Central Indo Pacific realm was more diverse and for Simpson diversity (q=2). Comparing Figs. 3b and 5b, we see that the samplesize- and coverage-based curves for q ¼ 0 and q ¼ 1 exhibit consistent diversity orderings between the two treatments. However, for q ¼ 2, the sample-size-based curves do not intersect (Fig. 3b), but the coverage-based curves have two crossing points (Fig. 5b). See Discussion for more comparisons of the two types of curves.

Step 1: Compare sample-size-based sampling curves up to a base sample size (Fig. 3).—We first constructed, for each of the two treatments, the integrated sample-sizebased rarefaction and extrapolation curves for Hill





Step 3: After calculating the coverage-based diversities of the 11 assemblages for q=0 , q=1, and the coverage-based sampling curves were compared up to a ‘‘base coverage’’, which is the lowest coverage for doubled reference sample sizes or the maximum coverage for reference samples, in the case of this project the base sample coverage was %.

, we compare the coverage-based diversities of the two treatments for q ¼ 0 (left panel), q ¼ 1 (middle panel), and q¼2 (right panel) up to the coverage of 96%. This is our ‘‘base coverage’’ (the lowest coverage for doubled reference sample sizes or the maximum coverage for reference samples, whichever is larger)

Since the two confidence bands do not intersect for species richness (q ¼ 0) if coverage exceeds 50% (Fig. 5b, left panel), species richness in the logged treatment is significantly higher than in

the girdled treatment for any standardized sample coverage between 50% and 96%. For Shannon diversity (q ¼ 1), the logged treatment is more diverse, but the confidence bands overlap. For Simpson diversity (q ¼ 2), when coverage is less than 70%, both treatments have almost the same diversity, but when coverage is greater than 70%, the Simpson diversity for the girdled treatment is slightly higher. Comparing Figs. 3b and 5b, we see that the sample

Revelante NW, Williams WT, Bunt JS (1982) Temporal and spatial distributions of diatoms, dinoflagellates and Trichodesmium in waters of the Great Barrier Reef. J Exp Mar Biol Ecol 63:27–45

Sano M (2000) Stability of reef fish assemblages: responses to coral recovery after catastrophic predation by Acanthaster planci. Mar Ecol Prog Ser 198:121–130