BIOTROPICA
THE JOURNAL OF THE ASSOCIATION FOR TROPICAL BIOLOGY AND CONSERVATION

A. TITLE PAGE

Jones, Rutherford, Deacon, Phillip, Magurran

FRESHWATER FISH BIODIVERSITY ESTIMATION

Title: **Quantifying regional biodiversity in the tropics: a case study of freshwater fish in Trinidad and Tobago.**

Authors:

Faith AM Jones[1]*, Mike G Rutherford[2], Amy E Deacon[2], Dawn AT Phillip[2,3], Anne E Magurran [1]

1. Centre for Biological Diversity, School of Biology, University of St Andrews, St Andrews, UK

2. Department of Life Sciences, The University of the West Indies, St Augustine, Trinidad and Tobago

3. Deceased 28 October 2017

*corresponding author (faj@st-andrews.ac.uk)

Received_____; revision accepted____.

In memory of Dr Dawn A T Phillip (1965 -2017)

B. ABSTRACT PAGE (Page 1)

Extinction rates are predicted to accelerate during the Anthropocene. Quantifying and mitigating these extinctions demands robust data on distributions of species and the diversity of taxa in regional biotas. However, many assemblages, particularly those in the tropics, are poorly characterized. Targeted surveys and historical museum collections are increasingly being used to meet the urgent need for accurate information, but the extent to which these contrasting data sources support meaningful inferences about biodiversity change in regional assemblages remains unclear. Here we seek to elucidate uncertainty surrounding regional biodiversity estimates by evaluating the performance of these alternative methods in estimating the species richness and assemblage composition of the freshwater fish of Trinidad and Tobago.

We compared estimates of regional species richness derived from two freshwater fish datasets: a targeted two year survey of Trinidad and Tobago rivers and historical museum collection records submitted to The University of the West Indies Zoology Museum. Richness was estimated using rarefaction and extrapolation, and assemblage composition was benchmarked against a recent literature review. Both datasets provided similar estimates of regional freshwater fish species richness (50 and 46 species, respectively), with a large overlap (85%) in species identities. Regional species richness estimates based on survey and museum data are thus comparable, and consistent in the species they include. Our results suggest that museum collection data are a viable option for setting reliable baselines in many tropical systems, thereby widening options for meaningful monitoring and evaluation of temporal trends.

C. KEY WORDS

1. Key words: Provide up to eight key words after the abstract, separated by a semi-colon (;). Key words should be in English (with the exception of taxonomic information) and listed alphabetically.

Assemblage composition; museum collections; species richness; neotropics; rarefaction; extrapolation

ALTHOUGH THERE IS GENERAL AGREEMENT THAT WE HAVE ENTERED THE ANTHROPOCENE, AN ERA LIKELY TO BE CHARACTERISED BY MASS EXTINCTIONS (Barnosky et al., 2011; Dirzo & Raven, 2003), there are substantial gaps in our understanding of biodiversity change, particularly at regional scales (McGill, Dornelas, Gotelli, & Magurran, 2015), and considerable uncertainty about extinction rates (Ceballos et al., 2015). Many assemblages, notably those in the tropics, are poorly characterised (Coddington, Agnarsson, Miller, Kunter, & Hormiga, 2009; Collen, Ram, Zamin, & McRae, 2008). Even in well-sampled areas many species are very rare, and are recorded in surveys only as singletons or “uniques” (Longino, Coddington, & Colwell, 2002). The presence of uniques in species accumulation curves is a strong indicator that unseen species are yet to be detected (Chao, 1984). One solution is to use statistical estimation approaches to deduce the number of unseen species in survey data (i.e. Chao & Jost, 2012; Gotelli & Colwell, 2001; Gotelli & Colwell, 2011).

Historical natural history museum records and herbarium collections are potential sources of data for biodiversity estimation, and are increasingly used to address ecological and conservation questions (Pyke & Ehrlich, 2010; Reznick, Baxter, & Endler, 1994). There are, though, concerns about possible biases in this type of data, particularly in terms of spatial representation and sampling bias (Fattorini, 2013; Guralnick & Van Cleve, 2005; Newbold, 2010).

The extent to which these different data sources provide meaningful inferences about biodiversity change in regional assemblages remains unclear. Survey data, on the one hand, may underestimate species richness to a greater extent than museum records because sampling is generally targeted at specific areas or habitats, or depends on methods which may incompletely record certain taxa (Guralnick & Van Cleve, 2005). For example, species that are known or suspected to be abundant in the sample area but are not easily recorded using the sampling methodology (Longino et al., 2002). On the other hand, while museums typically seek to maximise the range of specimens in the collection, they rarely set out to enumerate the species that co-occur in functioning ecosystems. Comprehensive species lists are accumulated over time, and often include transient taxa and misidentifications, so such lists are not necessarily an informative guide to the species actually present in an assemblage during a defined time period (Phillip et al., 2013).

Previous assessments of the relative utility of biodiversity quantifications from survey data and museum collections have focused on species richness rather than species identities (Guralnick & Van Cleve, 2005; Pyke & Ehrlich, 2010). However, biodiversity change can be substantially decoupled from species richness change when there is extensive turnover within assemblages (Dornelas et al., 2014; Hillebrand et al., 2017; Vellend et al., 2013, 2017). Accurate assessment of turnover (beta diversity), both spatial and temporal, is becoming increasingly important to understanding biodiversity change (Dornelas et al., 2014; McGill et al., 2015). There is consequently a need to recognize uncertainties and biases not only of species richness estimates, but also of species identities recorded within these contrasting datatypes. For example, previous research suggests that while museum records may provide useful estimates of richness, species identities may be biased towards rare species (Guralnick & Van Cleve, 2005; Pyke & Ehrlich, 2010). This is of particular concern if species lists derived from one sampling method will be used as baselines for further assessments using data collected with other methods.

Here we ask how conclusions about the biodiversity of freshwater fish in Trinidad and Tobago differ when based on a targeted survey versus a museum collection. First, we evaluate the performance of these alternative data sources when estimating the species richness of the freshwater fish fauna, and benchmark our results against a recent literature review. Secondly, we analyse the identities of species recorded by both methods to assess which species are absent, and to pinpoint possible biases in types of species detected.

Our initial expectations regarding biases in the datasets are as follows:

1. The museum data will contain more transient species than the survey data because the longer period of time covered by the museum data increases the chance of finding a species that subsequently becomes locally extinct.
2. There will be more species with specialized habitat requirements or narrow spatial distributions in the museum collection data than the sampling data. We expect this because of biases associated with museum collection data, specifically the "rare representation effect" where collectors target rare species (Guralnick & Van Cleve, 2005; Pyke & Ehrlich, 2010).
3. The majority of species missing from both datasets will be those that are narrowly distributed or habitat specialists, because these uncommon species are least likely to be noticed by collectors or sampled by systematic surveys.

METHODS

STUDY AREA

The country of Trinidad and Tobago is formed of two main islands lying to the northeast of Venezuela. Trinidad, the larger island, is 4820 km^2, and is only 11.3 km from Venezuela. Tobago is far smaller at 308 km^2, and sits 30.6 km from the coast of Venezuela. The climate of both islands is tropical, with a mean annual temperature of around 27°C, and a temperature range of around 17°C to 33°C. The islands support a variety of freshwater habitats. Streams in the north of Trinidad and in Tobago contain mostly clear, fast flowing water with firm substrate ranging from boulders to gravel. The more southern parts of Trinidad contain slower, more turbid streams, with substrates ranging from sand to mud.

Data Sources

Sampling was designed to provide useful data for conservation and management of the freshwater fish of Trinidad and Tobago. Ninety-one stream and river sites across the two islands were selected, representing all major drainages, biogeographic regions and river types. Each river had between one and three sampling locations. Sampling took place over two years (1997-1998), and 22 sites were sampled twice. Consistent sampling methods were used throughout, with small adjustments depending on stream type. Wherever possible, seine nets were used to block off sections of around 50m of river. A combination of methods including electrofishing (primarily in clear water), seine netting (in both clear and turbid water), and gill and trammel nets (particularly in larger rivers), were used to catch as many fish as possible in the blocked off sections. Species identities and their numerical abundances at each site were recorded before fish were returned to the stream at the point of capture.

The University of the West Indies Zoology Museum (UWIZM) is the de facto zoological collection for Trinidad & Tobago, and at the time of writing is one of the largest collections in the Caribbean. There are an estimated 70,000 specimens in the collections, the majority of which are local in origin. Although there was sporadic collecting of freshwater fish species from as early as 1936, the first significant fish collecting began in the mid-1960s and persisted through the rest of the 20th century. Few additions were made in the 2000s, but from 2010 onwards there were significant additions from work done by visiting researchers. The UWIZM data are open access, and available at https://doi.org/10.15468/m48ug8.

For our analysis, we use collection year as the collection unit of the museum data (Petersen & Meier, 2003). The nomenclature of the freshwater fish species in both the survey and the museum collection was also checked using the list of old and new species names provided by the species list and key of fish species (Phillip et al., 2013), ensuring that all names used in the final analysis were up to date and comparable.

ANALYSIS

Freshwater fish species lists, particularly those for islands, typically include species that are mostly restricted to freshwaters, and taxa that are either normally found in estuaries as well as fish that are predominately marine but occasionally move upstream. In addition these lists typically include anadromous and catadromous species. Here we follow Phillip et al. (2013)'s definition of freshwater fish, based on habitat preference and taxonomy. To identify these freshwater fish, we used a recent literature review that includes a comprehensive species list and key for fish species (Phillip et al., 2013). From this list, we selected only species that are considered by Phillip et al. (2013) as truly freshwater, not those that are usually regarded as marine or coastal species. We included transient species but not species that Phillip et al. (2013) considered misidentifications. DATP also submitted specimens and records to the museum between 1997 and 1998 as part of her survey. To avoid any confounding influence of these records on the museum collection data results, we removed all samples collected by DATP in 1997 and 1998.

To estimate freshwater fish species richness in Trinidad and Tobago, we used rarefaction and extrapolation curves computed by the 'iNEXT' R package (Chao et al., 2014). Extrapolation enables the user to estimate the number of species that would be detected if sampling was increased to include an additional number of individuals or sampling units. In individual based rarefaction, individuals should be sampled at random (Colwell et al., 2012), an expectation that museum data (and most ecological surveys) will not satisfy. But sample-based incidence data need only be representative of the area surveyed, including spatial heterogeneity (Chao & Colwell, 2017; Colwell, Mao, & Chang, 2004). Nonetheless, rarefaction and extrapolation has been shown to be a robust and informative method with different types of data (e.g. phylogenetic diversity (Chao et al., 2015; Hsieh, Ma, & Chao, 2016b) and distributions of stone tools in Pleistocene North America (Buchanan et al., 2017).

Also, note that the estimate attained from extrapolation is exactly the same as the non-parametric Chao2 estimate. For both datasets, we used the "incidence_freq" datatype option, that is sample-based rarefaction rather than individual based rarefaction. For sample-based rarefaction sampling units need only be representative of the sampling area, which is a less stringent assumption than for individual based rarefaction (Colwell et al., 2004). We therefore chose sample-based rarefaction rather than individual based rarefaction in both cases. For the Museum data, we used the year in which an acquisition was recorded as its sample id. We benchmarked the estimated species richness numbers against the number provided by a comprehensive species list collated using all available fish records and expert knowledge of the Trinidad and Tobago freshwater fish fauna (Phillip et al., 2013)

To further understand whether the survey dataset and the museum collection dataset differ in the types of fish they represent, we categorized each fish species by status (i.e. native/non-native), by habitat specificity, and by how widely it was distributed across Trinidad and Tobago, using information in Phillip et al. (2013) and FishBase (fishbase.org) - see Table 1 and Table S1. We compared the distribution of characteristics of the species observed in both the survey and museum datasets against the results of a null model (Fig. S1). The assumption of the null model was that each species had equal probability of being recorded, as long as it is found, or has been found, in the rivers of Trinidad and Tobago. For each iteration of the null model, 39 species (the number of observed species in the Museum data) were randomly selected from the list of 65 species that are likely to be present in Trinidad and Tobago according to Phillip et al. (2013). We recorded the native status, distribution, and habitat specificity of each of the randomly selected species, and then proceeded with the next iteration. The model had 1000 iterations. We then calculated the quantiles of the observed numbers of fish in the survey and museum data for each category in relation to the null model results.

RESULTS

ESTIMATED RICHNESS

Visual inspection of the observed species richness accumulation curve for the survey data (Fig. 1a), suggests an asymptote is close. Although there are far more records overall in the survey data than the museum data (Table 2), most species are found at only a few sites (under 20) – a typical pattern in ecological surveys (Fig. 1c).

In contrast to the survey data results, the museum data accumulation curve does not support an asymptote close to the 39 species recorded (Fig. 1b). Collection effort is extremely variable in the museum data, with over 200 records submitted for one year in the 1990s and fewer than 100 for most other years (Fig. 1d). There is, however, no noticeable increase in new species during the period of increased specimen submissions (Fig. 1b). In addition, both data collection methods provide samples that are close to completely representative (Fig. 1e & 1f).

The iNEXT extrapolations estimated were within 10% of each other (50 species for the survey data (Fig. 1g), 46 for the museum data (Fig. 1h)), and they both lie well within each other's upper and lower 95% confidence intervals (Table 2). The survey data had higher uncertainty around this estimate, with the upper 95% richness estimated as 130 species as opposed to the 68 estimated from the museum data. The range of estimates predicted by both data types included the 65 species reported by the comprehensive key and species list (Phillip et al., 2013) (note the 66 quoted in the text of (Phillip et al., 2013) is a miscount of the true number listed in the table of species).

Assemblage composition

Fewer species are missing from the museum data but recorded in the survey data (4) than recorded in the survey but missing from the museum data (6) (Fig.2). No transient species were recorded in either dataset (Table 3), and the majority of species in both datasets were native. A high proportion of species missed by both data collection methods either were data deficient, transient, narrowly distributed or habitat specialists (Table 3).

Contrary to our expectation, there were no biases evident between types of fish recorded in the survey and museum data (Table 3; Fig. 3). Both underestimated the number of species thought to be present in Trinidad and Tobago, but the fraction of native species was higher in both cases than in the overall list provided by Phillip et a*l*. (2013). Both datasets also included more intermediately or widely distributed species than this overall list, although the difference was more marked in the museum data than the survey data. There are also more habitat generalists in the observed data (both methods) than expected if they were a random draw from the overall list. This difference, however, is less pronounced because the number of habitat generalists in both surveys fall within the 95% quantiles of the null model.

DISCUSSION

Despite the two orders of magnitude fewer records contained in the museum data than the targeted survey data, both datasets provided comparable estimates of regional freshwater fish species richness in Trinidad and Tobago. The richness estimates of the museum and survey data were within 10% of each other (50 species and 46 species, respectively), and there was a large overlap (85%) of species identified. Both estimates fall 20% below the maximum number (65) of species potentially present according to the exhaustive list (Phillip et al., 2013; Table 2), but the upper confidence intervals of the estimates are inclusive of this maximum number of potential species.

We expected differences in the composition of species observed in the two contrasting datasets because of biases in the collection methods of the museum data. For example, sampling in historical museum collections generally occurs *ad hoc* by a variety of uncoordinated collectors, typically leading to an overrepresentation of easily accessible areas and centres of population (Engemann et al., 2015; Guralnick & Van Cleve, 2005; Soberón, Llorente, & Oñate, 2000; Tobler, Honorio, Janovec, & Reynel, 2007). Another bias is the "rare representation" effect: the tendency for collectors to favour unusual species, combined with longer collection times, giving a greater likelihood of finding species outside of their usual ranges (Guralnick & Van Cleve, 2005; Pyke & Ehrlich, 2010). The rare representation effect could cause overestimations of species richness, which in turn might inflate the importance of transient species that do not contribute to ecosystem processes. Contrary to our expectations, we found a striking similarity between the identities of the species recorded in the survey and museum data (Figs 1 & 3), suggesting these biases do not strongly influence regional species richness estimates in these data. The majority (85%) of species were recorded in both datasets. In addition, there was no indication of biases in types of species recorded; the museum collection data did not contain more transient species, nor habitat specialists or narrowly distributed species, than the survey data.

Our results suggest that, although collection methods differ considerably between datasets, survey and museum data can provide comparable estimations of the regional assemblage species composition. The substantial overlap in species present in both datasets is particularly notable because the dissimilarity between samples is inflated by incomplete species lists (Chao, Chazdon, Colwell, & Shen, 2005). Consequently, historical museum collection data are potentially useful for analysing other aspects of biodiversity change in addition to richness. Rates of turnover of species identity within assemblages, for instance, could be assessed with species lists. Rates of turnover are variable and driven by a complex collection of biotic and abiotic factors (Korhonen, Soininen, & Hillebrand, 2010), and warrant more analysis. Datasets such as the collections held at The University of the West Indies Zoology Museum, Trinidad, could serve as a baseline for furthering our understanding of turnover within communities. Within the Caribbean region, for instance, collections similar to those held by The University of the West Indies Zoology Museum, Trinidad, include those held at The National Zoological Collection of Suriname (NZCS) and The Museo Nacional de Historia Natural ¨prof. Eugenio de Jesus Marcano¨ in Santo Domingo, Dominican Republic. More widely, there are similar museums with extensive collections that could be used to form the basis of species lists in Costa Rica, Cuba, Venezuela, Colombia, Panama and Nicaragua. There are also increasing possibilities for searching for and combining collections from multiple sources as more museum collection data are uploaded onto online repositories like the Global Biodiversity Information Facility (GBIF), meaning collections held outside of tropical regions can also be harnessed for creating baseline species lists.

Surveys provide robust data on species distributions and abundance, and are generally suitable for a wider variety of analyses than museum data. For example, the combination of species identity and relative abundance values of systematic survey data mean diversity metrics such as Hill numbers (which include forms of Shannon and Simpson diversity measures) can be calculated (Hill, 1973). These estimates allow the almost unbiased "effective" number of frequent species within assemblages to be estimated (Hsieh, Ma, & Chao, 2016a). However, surveys are not practical in many cases. Undertaking surveys can be expensive and requires good access to expertise and sites. The survey we used in this analysis took place over two years, and involved many hours of preparation and field work. Even in relatively well sampled sites, a short period of sampling activity does not often come close to the actual number of species in an area (Fattorini, 2013). This is a particular problem in tropical regions, where there is a substantial need for data. Alternative data gathering exercises, namely intensive local sampling areas (i.e. Bouchet, Lozouet, Maestrati, & Heros, 2002; Brown et al., 2018; Longino et al., 2002) could also be useful, but these sampling endeavours also require extremely high levels of expertise and investment, which are often unavailable, and are not practical on a regional scale. In these cases, museum and other historical natural history collections provide a useful resource for estimating regional species richness. This is not to say that historical museum data can or should replace systematic survey. For instance, an aspect of biodiversity change that may strongly affect ecosystem functioning is reordering of species abundances with assemblages (Jones, Ripplinger, & Collins, 2017). To what extent such reordering of community structure, in particularly whether dominant species are changing identity, requires representative relative abundance data, which cannot be extracted from *ad hoc* museum collections.

While both datasets investigated in this study gave similar estimates of species richness and assemblage compositions, there was substantial divergence between their estimates and that of a recent literature review and key (Phillip et al., 2013; Table 3). The species missed from both datasets tended to be narrowly distributed habitat specialists or recent additions to the Trinidad and Tobago freshwater fauna, and may include some species that were presumed native but may not be currently present in the region. Such biases are extremely common in ecological assemblage data (Longino et al., 2002), with most undescribed species believed to be narrowly distributed and uncommon within their home ranges (Pimm et al., 2014). These biases raise the question of whether both our empirical datasets underestimate species richness, or whether the exhaustive list compiled from a literature search is an overestimate. This is an important consideration, because how much emphasis is given to the most difficult to detect species in an assemblage heavily influences estimated extinction and turnover rates. Recently detected species may go extinct before or just after their discovery (Barnosky et al., 2011; Lees & Pimm, 2015), particularly if they are transient species (Magurran & Henderson, 2003) or have restricted distributions (Pimm et al., 2014).

Concluding remarks

Uncertainty around biodiversity levels and distribution hinders our understanding of key biodiversity statistics and consequently our ability to make informed conservation decisions (Pimm et al., 2014). Understanding the information gaps around biodiversity knowledge is essential for progression of the field (Hortal et al., 2015). In our analysis we demonstrated that both historical museum collection data and survey data can provide useful regional species richness estimates to use as baselines for assessing biodiversity change. Both datasets also provided comparable estimates of the identities of species within the assemblage, as they detected all but the transient or very difficult to detect species. Most assemblages display similar species abundance distributions, characterised by both common and rare species (McGill et al., 2007) and often include both "core" and "transient" species (Magurran & Henderson, 2003; Taylor, Evans, White, & Hurlbert, 2018). Our results suggest that the majority of a region's "core" species are detected by both museum data and survey data to similar extents. Consequently, species lists for assessing turnover within tropical regions, and amongst these "core taxa", could be compiled from existing historical museum collections where suitable systematic survey data are unavailable. This would provide opportunities for monitoring and understanding biodiversity change within tropical regions that otherwise lack appropriate baseline data.

Then again, it is difficult to verify which of the fish species potentially in Trinidad and Tobago are actually present in the region at a given time. This uncertainty needs to be taken into account in baseline estimates of regional species richness and turnover/extinction analyses.

Based on our results, and with appropriate caveats, we therefore recommend increased use of historical museum collections, particularly those containing tropical data, in assessments of regional biodiversity. These data are more readily available than intensive systematic survey data in many parts of the world, and assemblage composition within such collections can be sufficiently unbiased as to serve as useful baselines for assessing temporal turnover of species identities. By harnessing their full potential, we can provide a useful source of biodiversity information to help bridge the knowledge gap between temperate and tropical systems.

E. ACKNOWLEDGMENTS

Dawn Phillip's PhD was funded by the Darwin Initiative, UK. AEM thanks the European Research Council (AdG BioTIME 250189 and PoC BioCHANGE 727440), and Dr Mary Alkins-Koo and Professor Indar Ramnarine for their support of Dr Phillip during her doctoral research. We are grateful to Dr Charles Paxton for advice on statistical methods. We are indebted to Professor Robert Colwell and an anonymous reviewer, and to the editor, for insightful and constructive comments on a previous version of this manuscript. Finally, we wish to thank Ms Ada Fontrodona Eslava for kindly translating out abstract into Spanish.

F. DATA ACCESSIBILITY

The Phillip survey data are available at https://doi.org/10.17630/e60bce64-2926-42d4-a69f-91e36ade4629. The UWIZM data are open access, and available at https://doi.org/10.15468/m48ug8

G. LITERATURE CITED

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., … Ferrer, E. A. (2011). Has the Earth’s sixth mass extinction already arrived? *Nature*, *471*(7336), 51–57. https://doi.org/10.1038/nature09678

Bouchet, P., Lozouet, P., Maestrati, P., & Heros, V. (2002). Assessing the magnitude of species richness in tropical marine environments : exceptionally high numbers of molluscs at a New Caledonia. *Biological Journal of the Linnean Society*, *75*, 421–436. https://doi.org/10.1046/j.1095-8312.2002.00052.x

Brown, B. V., Borkent, A., Adler, P. H., Amorim, D. de S., Barber, K., Bickel, D., … Zumbado, M. A. (2018). Comprehensive inventory of true flies (Diptera) at a tropical site. *Communications Biology*, *1*(1), 21. https://doi.org/10.1038/s42003-018-0022-x

Buchanan, B., Chao, A., Chiu, C. H., Colwell, R. K., O’Brien, M. J., Werner, A., & Eren, M. I. (2017). Environment-induced changes in selective constraints on social learning during the peopling of the Americas. *Scientific Reports*, *7*(March), 1–12. https://doi.org/10.1038/srep44431

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human–induced species losses: entering the sixth mass extinction. *Sciences Advances*, *1*(e1400253), 1–5. https://doi.org/10.1126/sciadv.1400253

Chao, A. (1984). Nonparametric Estimation of the Number of Classes in a Population. *Scandinavian Journal of Statistics*, *11*(4), 265–270.

Chao, A., Chazdon, R. L., Colwell, R. K., & Shen, T.-J. (2005). A new statistical approach for assessing similarity of species composition with incidence and abundance data.

Ecology Letters, *8*, 148–159. https://doi.org/10.1111/j.1461-0248.2004.00707.x

Chao, A., Chiu, C. H., Hsieh, T. C., Davis, T., Nipperess, D. A., & Faith, D. P. (2015). Rarefaction and extrapolation of phylogenetic diversity. *Methods in Ecology and Evolution*, *6*(4), 380–388. https://doi.org/10.1111/2041-210X.12247

Chao, A., & Colwell, R. K. (2017). Thirty years of progeny from Chao's inequality: Estimating and comparing richness with incidence data and incomplete sampling. *Sort*, *41*(1), 3–54.

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs*, *84*(1), 45–67. https://doi.org/10.1890/13-0133.1

Chao, A., & Jost, L. (2012). Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology*, *93*(12), 93–95. https://doi.org/10.1890/11-1952.1

Coddington, J. A., Agnarsson, I., Miller, J. A., Kunter, M., & Hormiga, G. (2009). Undersampling bias: the null hypothesis for singleton species in tropical arthropod surveys. *Journal of Animal Ecology*, *78*, 573–584. https://doi.org/10.1111/j.1365-2656.2007.0

Collen, B., Ram, M., Zamin, T., & McRae, L. (2008). The Tropical Biodiversity Data Gap: Addressing Disparity in Global Monitoring. *Tropical Conservation Science*, *1*(2), 75–88. https://doi.org/10.1177/194008290800100202

Colwell, R. K., Chao, A., Gotelli, N. J., Lin, S. Y., Mao, C. X., Chazdon, R. L., & Longino, J. T. (2012). Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology*, *5*(1), 3–21. https://doi.org/10.1093/jpe/rtr044

Colwell, R. K., Mao, C. X., & Chang, J. (2004). Interpolating, Extrapolating, and Comparing Incidence-Based Species Accumulation Curves. *Ecology*, *85*(10), 2717–2727. https://doi.org/10.1890/03-0557

Dirzo, R., & Raven, P. H. (2003). Global State of Biodiversity and Loss. *Annual Review of Environment and Resources*, *28*(1), 137–167. https://doi.org/10.1146/annurev.energy.28.050302.105532

Dornelas, M., Gotelli, N. J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C., & Magurran, A. E. (2014). Assemblage time series reveal biodiversity change but not systematic loss. *Science*, *344*(6181), 296–299. https://doi.org/10.1126/science.1248484

Engemann, K., Enquist, B. J., Sandel, B., Boyle, B., Jørgensen, P. M., Morueta-Holme, N., … Svenning, J. C. (2015). Limited sampling hampers "big data" estimation of species richness in a tropical biodiversity hotspot. *Ecology and Evolution*, *5*(3), 807–820. https://doi.org/10.1002/ece3.1405

Fattorini, S. (2013). Regional Insect Inventories Require Long Time, Extensive Spatial Sampling and Good Will. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0062118

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, *4*, 379–391. https://doi.org/10.1046/j.1461-0248.2001.00230.x

Gotelli, N. J., & Colwell, R. K. (2011). Estimating species richness. In A. E. Magurran & B. J. McGill (Eds.), *Biological diversity; fronteirs in measurement and assessment* (pp.

39–54). Oxford University Press. https://doi.org/10.2307/3547060

GURALNICK, R., & VAN CLEVE, J. (2005). Strengths and weaknesses of museum and national survey data sets for predicting regional species richness: Comparative and combined approaches. *Diversity and Distributions*, *11*(4), 349–359. https://doi.org/10.1111/j.1366-9516.2005.00164.x

HILL, M. O. (1973). Diversity and evenness: a unifying notation and its consequences. *Ecology*, *54*(2), 427–432. https://doi.org/10.2307/1934352

HILLEBRAND, H., BLASIUS, B., BORER, E. T., CHASE, J. M., DOWNING, J., ERIKSSON, B. K., … RYABOV, A. B. (2017). Biodiversity change is uncoupled from species richness trends. *Journal of Applied Ecology*, *55*(March), 1–16. https://doi.org/10.1111/ijlh.12426

HORTAL, J., DE BELLO, F., DINIZ-FILHO, J. A. F., LEWINSOHN, T. M., LOBO, J. M., & LADLE, R. J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

HSIEH, T. C., MA, K. H., & CHAO, A. (2016a). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, *7*(12), 1451–1456. https://doi.org/10.1111/2041-210X.12613

HSIEH, T. C., MA, K. H., & CHAO, A. (2016b). iNEXT: iNterpolation and EXTrapolation for species diversity. Retrieved from url:%0A http://chao.stat.nthu.edu.tw/blog/software-download/.

JONES, S. K., RIPPLINGER, J., & COLLINS, S. L. (2017). Species reordering, not changes in richness, drives long-term dynamics in grassland communities. *Ecology Letters*, *20*(12), 1556–1565. https://doi.org/10.1111/ele.12864

Korhonen, J. J., Soininen, J., & Hillebrand, H. (2010). A quantitative analysis of temporal turnover in aquatic species assemblages across ecosystems. *Ecology*, *91*(2), 508–517. https://doi.org/10.1890/09-0392.1

Lees, A. C., & Pimm, S. L. (2015). Species, extinct before we know them? *Current Biology*, *25*(5), R177–R180. https://doi.org/10.1016/j.cub.2014.12.017

Longino, J. T., Coddington, J., & Colwell, R. K. (2002). The Ant Fauna of a Tropical Rain Forest : Estimating Species Richness Three Different Ways. *Ecology*, *83*(3), 689–702. https://doi.org/10.2307/3071874

Magurran, A. E., & Henderson, P. A. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature*, *422*(6933), 714–716. https://doi.org/10.1038/nature01547

McGill, B. J., Dornelas, M., Gotelli, N. J., & Magurran, A. E. (2015). Fifteen forms of biodiversity trend in the anthropocene. *Trends in Ecology and Evolution*, *30*(2), 104. https://doi.org/10.1016/j.tree.2014.11.006

McGill, B. J., Etienne, R. S., Gray, J. S., Alonso, D., Anderson, M. J., Benecha, H. K., … White, E. P. (2007). Species abundance distributions: Moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, *10*(10), 995–1015. https://doi.org/10.1111/j.1461-0248.2007.01094.x

Newbold, T. (2010). Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography*, *34*(1), 3–22. https://doi.org/10.1177/0309133309355630

Petersen, F. T., & Meier, R. (2003). Testing species richness estimation methods on single sample collection data using the Danish Diptera. *Biodiversity and Conservation*, 687–

702.

Phillip, D. A. T. (1998). *Biodiversity of the freshwater fishes of Trinidad and Tobago.* University of St Andrews, Scotland, UK. Retrieved from http://research-repository.st-andrews.ac.uk/handle/10023/2832

Phillip, D. A. T., Taphorn, D. C., Holm, E., Gilliam, J. F., Lamphere, B. A., & López-Fernández, H. (2013). *Annotated list and key to the stream fishes of Trinidad & Tobago*. *Zootaxa* (Vol. 3711). https://doi.org/10.11646/zootaxa.3711.1.1

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., … Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, *344*(6187). https://doi.org/10.1126/science.1246752

Pyke, G. H., & Ehrlich, P. R. (2010). Biological collections and ecological/environmental research: A review, some observations and a look to the future. *Biological Reviews*, *85*(2), 247–266. https://doi.org/10.1111/j.1469-185X.2009.00098.x

Reznick, D., Baxter, R. J., & Endler, J. (1994). Long-Term Studies of Tropical Stream Fish Communities : The Use of Field Notes and Museum Collections to Reconstruct Communities of the Past. *American Zoologist*, *34*(3), 452–462.

Soberón, J. M., Llorente, J. B., & Oñate, L. (2000). The use of specimen-label databases for conservation purposes: an example using Mexican Papilionid and Pierid butterflies. *Biodiversity and Conservation*, *9*, 1441–1466.

Taylor, S. S., Evans, B. S., White, E. P., & Hurlbert, A. H. (2018). The prevalence and impact of transient species in ecological communities. *Ecology*, *0*(0), 1–11. https://doi.org/10.1002/ecy.2398

TOBLER, M., HONORIO, E., JANOVEC, J., & REYNEL, C. (2007). Implications of collection patterns of botanical specimens on their usefulness for conservation planning: An example of two neotropical plant families (Moraceae and Myristicaceae) in Peru. *Biodiversity and Conservation*, *16*(3), 659–677. https://doi.org/10.1007/s10531-005-3373-9

VELLEND, M., BAETEN, L., MYERS-SMITH, I. H., ELMENDORF, S. C., BEAUSÉJOUR, R., BROWN, C. D., … WIPF, S. (2013). Global meta-analysis reveals no net change in local-scale plant biodiversity over time. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(48), 19456–19459. https://doi.org/10.1073/pnas.1312779110

VELLEND, M., DORNELAS, M., BAETEN, L., BEAUSEJOUR, R., BROWN, C., DE FRENNE, P., … SIEVERS, C. (2017). Estimates of local biodiversity change over time stand up to scrutiny. *Ecology*, *98*, 583–590. https://doi.org/10.1101/062133

Tables

Table 1. Category descriptions for assigning species characteristics. Information was extracted from (Phillip et al., 2013). Any fish described as "mistake" were removed from the analysis during data preparation.

Designation	Description
Status	
No data	No data on this characteristic in the fish key
Introduced	Species colonised from a human introduction
Mistake	Misidentifications
Presumed native	Presumed native to Trinidad and Tobago
Recent Colonist	Natural colonists from the Orinoco River
Transient	Species not recorded in the last 2 to 3 surveys. They are natural colonists from the Orinoco River that did not become established
Habitat Specificity	
No data	No data on this characteristic in the fish key
Specialist	Lives in only one water type, i.e. clear and fast flowing
Generalist	Can live in different water types, i.e. clear fast flowing water and turbid water
Distribution	
No data	No data on this characteristic in the fish key
Narrow	Only found in a few sites

Intermediate	Either found in a subsection of Trinidad that is more than a few streams or fish described as “widely distributed” in a subsection of Trinidad
Wide	Found in most of Trinidad, or found in both Trinidad and Tobago, or described as "widely distributed"

Table 2. A breakdown of the numbers of acquisition records uniques (species only recorded once), duplicates (species recorded twice), and the observed number of species in the sampling and museum freshwater fish, as well as the number of freshwater fish estimated to be extant in Trinidad and Tobago according to rarefaction and extrapolation using iNEXT. These species richness estimate are exactly that of the non-parametric Chao 2 estimate.

Dataset type	**Acquisitions**	**Sampling Units**	**Uniques**	**Duplicates**	**Species Observed**	**Species Richness estimate**	**Lower bound**	**Upper bound**
Survey	21153	56	4	3	38	50 (+/- 17)	40	131
Museum	785	30	2	3	39	46(+/- 6)	40	68

Table 3. A breakdown of the status, distribution and habitat preference characteristics of all freshwater fish species in Trinidad and Tobago as stated in (Phillip et al., 2013). A further breakdown of the characteristics of the fish found in the survey and museum datasets is also included, and the quantiles of these values in relation to the null model results. Finally, we include a breakdown of the characteristics of the fish found in neither the Survey nor the Museum data.

		All Species	Survey	Museum	Survey Quantile	Museum Quantile	Not found
Status	No data	3	0	0	0.18	0.18	3
	Introduced	5	2	2	0.42	0.42	3
	Native	53	33	34	1.00	1.00	16
	Recent Colonist	4	3	3	0.92	0.92	0
Distribution	No data	11	0	0	0.00	0.00	11
	Narrow	13	6	4	0.34	0.05	6
	Intermediate	24	19	22	1.00	1.00	2
	Wide	17	13	13	0.99	0.99	3
Habitat	No data	2	0	0	0.00	0.00	2
	Specialist	23	14	14	0.84	0.84	7
	Generalist	40	24	25	0.94	1.00	13

Figures

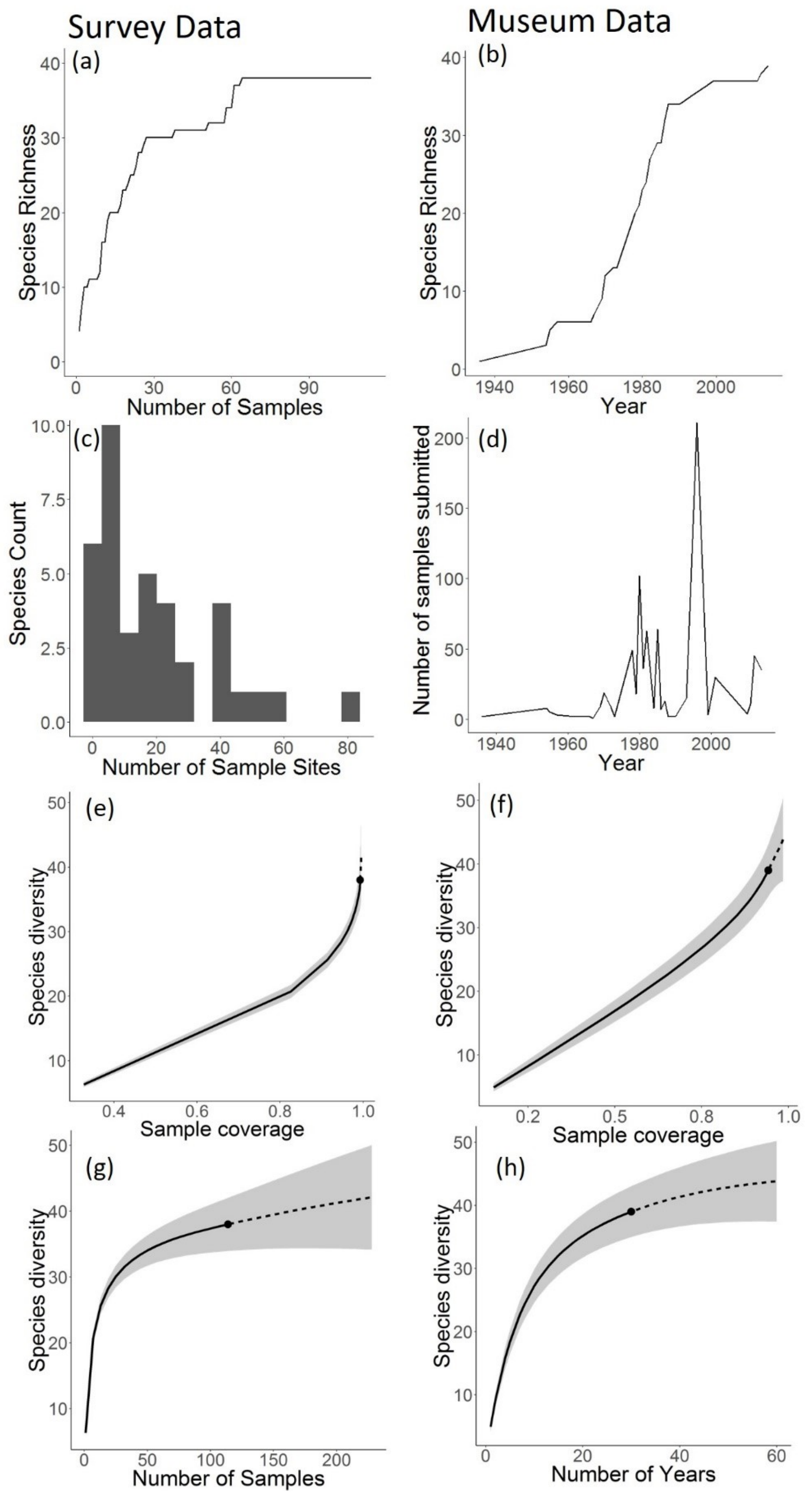
Survey Data
Museum Data
(a)
Species Richness
Number of Samples
(b)
Species Richness
Year
(c)
Species Count
Number of Sample Sites
(d)
Number of samples submitted
Year
(e)
Species diversity
Sample coverage
(f)
Species diversity
Sample coverage
(g)
Species diversity
Number of Samples
(h)
Species diversity
Number of Years

Figure 1. Plots of the Trinidad and Tobago freshwater fish targeted survey data and museum collection data. Plot (a) is the accumulation of species richness as new sites were added to the survey data in terms of the actual temporal sequence of data collection. Plot (b) shows the accumulation of species richness in the museum collections through time. Plot (c) shows the frequency of species found in multiple sites, and plot (d) shows the unequal distribution of sample submissions to the museum collection over time. Plots (e) and (f) show the coverage-based extrapolation for the survey and museum data respectively, and (g) and (h) show the estimated species richness of the survey and museum data, respectively, using the iNEXT sample-based extrapolation. The grey ribbon represents the 95% Confidence Intervals of the estimates.

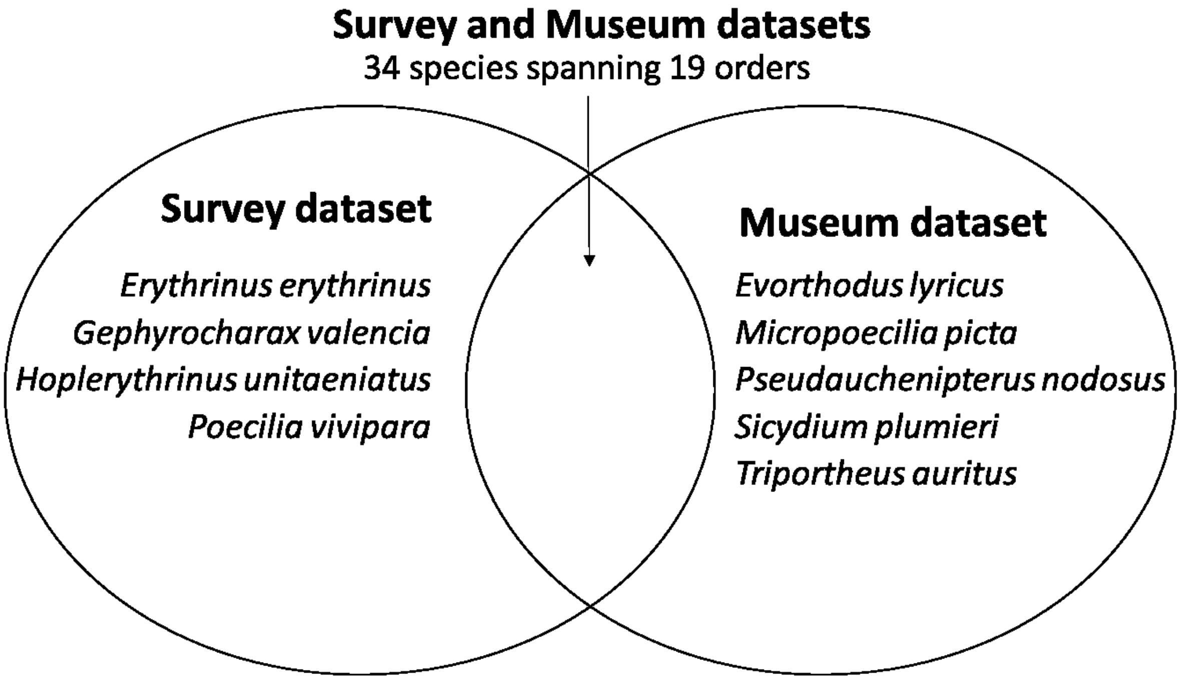

Figure 2. A breakdown of which species were recorded only in the survey data and only in the museum data. The majority (34) species were recorded in both datasets. For a complete list of which species where found in each dataset see Table S1.

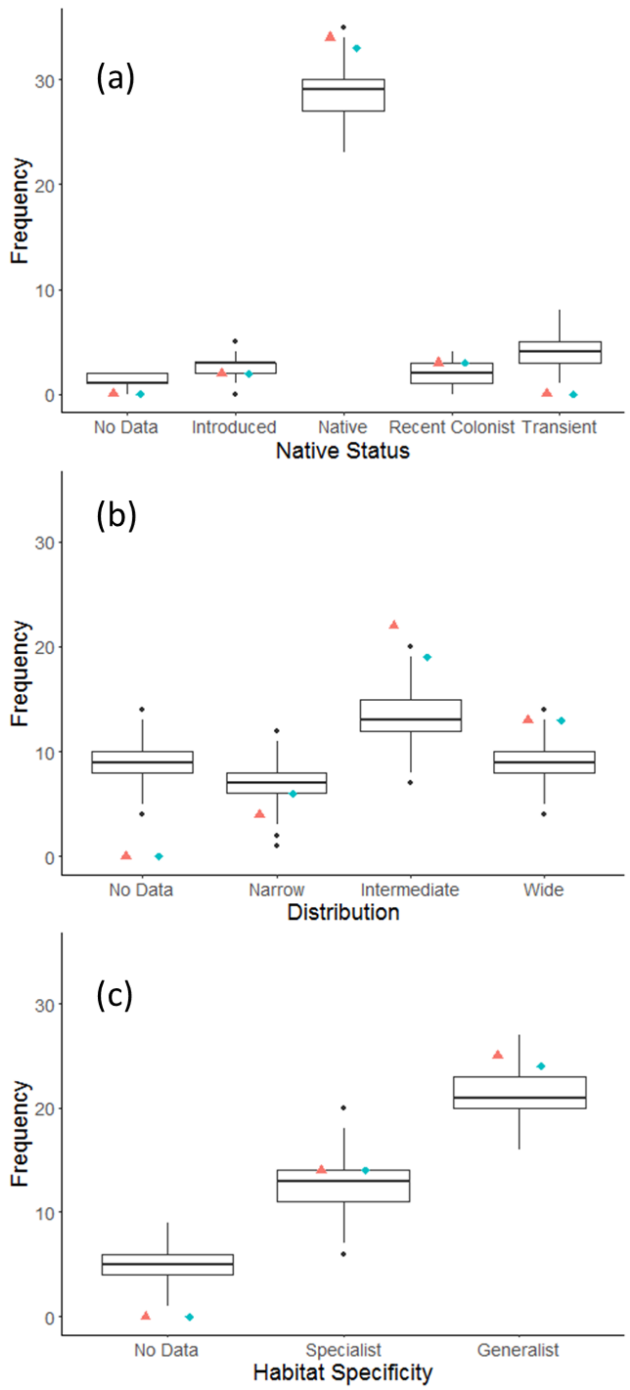

Figure 3. The native status (a), distribution (b) and habitat specificity (c) of 39 fish randomly selected by a null model with 1000 iterations (black boxplots), compared to the observed habitat specificity of species found in the survey data (red triangles) and museum data (blue diamonds). Box plots show medians, upper and lower quantiles and outliers. A violin plot showing the observed values and the sampling distribution of the model can be found in Fig. S2.

In memory of Dr Dawn A T Phillip (1965 -2017)